รายงานวิจัยฉบับสมบูรณ์

โครงการ การคัดเลือกปริภูมิย่อยความละเอียดสูงสุดยิ่งยวดที่ดีที่สุดในการรู้จำใบหน้าและเป้าหมายอัตโนมัติ: วิธีปฏิบัติด้วยคลาสของการวิเคราะห์มุขสำคัญระบุทิศทาง

โดย

ผู้ช่วยศาสตราจารย์ ดร.วิทยากร อัศดรวิเศษ

เดือน กรกฎาคม ปี 2554

สัญญาเลขที่ MRG5080427

รายงานวิจัยฉบับสมบูรณ์

โครงการ การคัดเลือกปริภูมิย่อยความละเอียดสูงสุดยิ่งยวดที่ดีที่สุดในการรู้จำใบหน้าและเป้าหมายอัตโนมัติ: วิธีปฏิบัติด้วยคลาสของการวิเคราะห์มุขสำคัญระบุทิศทาง

ผู้วิจัย ผู้ช่วยศาสตราจารย์ ดร.วิทยากร อัศดรวิเศษ
สังกัด จุฬาลงกรณ์มหาวิทยาลัย

สนับสนุนโดยสำนักงานคณะกรรมการการอุดมศึกษา
และสำนักงานกองทุนสนับสนุนการวิจัย

บทคัดย่อ

วัตถุประสงค์ของโครงการวิจัยนี้ เพื่อศึกษาการปรับปรุงสมรรถนะของ *การวิเคราะห์ส่วนประกอบมุขสำคัญสองมิติ* (two-dimensional principal component analysis :2DPCA) และส่วนขยาย ทั้งในทางทฤษฎี และ การคำนวณเชิงเลข โดยเทคนิคการสร้างคืนความละเอียดสูงยวดยิ่ง และ การคัดเลือกตัวแปร การศึกษานี้จะเน้นการประยุกต์ใช้ในการรู้จำภาพใบหน้าและเป้าหมายทางทหาร โดยการใช้ฐานข้อมูลมาตรฐาน เช่น Yale, AR, ORL และ MSTAR 2DPCA ประสบความสำเร็จในการประยุกต์ใช้งานกับปัญหาการรู้จำรูปแบบสองมิติโดยเฉพาะการเข้าใจภาพ

ในการศึกษานี้ เราได้ผัน 2DPCA แบบระบุทิศจาก 2DPCA แนวขวาง นอกจากนั้น เราได้นำเสนอ 2DPCA ไขว้ ที่ได้จากการปรับปรุง เมทริกซ์ความแปรปรวนร่วมเกี่ยวของ PCA อย่างมีแบบแผน เรายังได้พบว่า ทั้ง 2DPCA แบบระบุทิศ และ ไขว้ ต่างก็มีความเป็นไปได้ทางชีวภาพของการมองเห็น ด้วยเหตุผลหลายประการที่สามารถอธิบายได้ ว่า 2DPCA น่าจะทำงานคล้ายกับระบบมองเห็นของเรา โดยเฉพาะมีการพบว่า เซลในจอประสาทตาและทางเดินของการมองเห็นของสัตว์ลี้ยงลูกด้วยนมส่วนใหญ่ มีลักษณะที่อ่อนไหวต่อทิศ ระบบจำแนกแบบหลายตัวได้ถูกนำเสนอเนื่องมาจากเหตุผลที่ว่า ปมประสาท โดยส่วนใหญ่ทำงานแบบไม่สมมาตร และ เป็นอิสระต่อกัน ดังนั้น วิธีจำแนกเชิงองค์คณะหลายวิธีจึงได้ถูกศึกษา เช่น วิธีปริภูมิสุ่ม (random subspace method) หรือ การรวมแบบคัดเลือกตัวบ่งต่าง เงื่อนไขที่ใช้ในการรวมแบบคัดเลือกตัวบ่งต่าง ได้แก่ ระยะห่าง *Kullback-Leibler*

อีกแง่หนึ่งที่รวมอยู่ในการศึกษาครั้งนี้ ได้แก่ การศึกษาถึงคุณสมบัติการเลือกสรรที่ขึ้นกับทิศของภาพธรรมชาติ ในที่นี้ เราอาจสร้าง เมทริกซ์ความแปรปรวนร่วมเกี่ยวไขว้สองมิติ จากเมทริกซ์ความแปรปรวนร่วมเกี่ยวหนึ่งมิติ โดยการใช้เวฟเลตระบุทิศ ด้วยวิธีนี้ เราจะสามารถสร้าง 2DPCA แบบใหม่ เพื่อแทนที่ 2DPCA แบบระบุทิศ และ ไขว้ ที่กล่าวมาก่อนหน้านี้ ผลของการศึกษาอย่างกว้างขวางของเราก่อให้เกิดปัญหาวิจัยเปิดมากมายที่น่าจะศึกษาต่อไปในอนาคต

Abstract

The aim of this project is to theoretically and numerically study the performance improvement of two-dimensional principal component analysis (2DPCA) and its variants by employing super-resolution techniques and variable selection methods. This study will mainly focus on face and automatic target recognition applications using standard databases as Yale, AR, ORL and MSTAR. 2DPCA has been successfully applied to various two-dimensional pattern recognition problems, especially in image understandings.

In this study, we derive directional 2DPCA from diagonal 2DPCA. Furthermore, we introduce crossed-2DPCA by modifying the PCA covariance matrix heuristically. We also investigate and find out that both directional and crossed-2DPCA are biological plausible. There are quite a number of explainable reasons that 2DPCA use be used to imitate our vision systems. In particular, cell in major vertebrate retina and visual pathways as well as our proposed 2DPCA are directionally selective. Multiple classifier systems has been proposed to use based on the fact that all ganglion are working irregularly and independently. This way, several ensemble-based classification methods have been investigated, such as random subspace method and feature selective combining. The criteria used for our feature selective combining is *Kullback-Leibler* distance.

Another issue that will be included in this study is to study the property of the directional selectivity of natural image. Here, the new 2D cross-covariance matrices can be constructed from 1-D covariance matrix using directional wavelet. This way, we can get new directional 2DPCA to replace the above directional and crossed-2DPCA. Our extensive study posed many open research problems that should be investigated in the future.

กิตติกรรมประกาศ

งานวิจัยฉบับนี้ ส่วนหนึ่งได้รับการสนับสนุนเงินทุนวิจัยภายใต้โครงการพัฒนาฯอาจารย์ใหม่ สกว. ร่วมกับ สำนักงานคณะกรรมการการอุดมศึกษา ในการทำการวิจัยตลอดโครงการ จนงานสำเร็จลุล่วงไปด้วยดี ผู้วิจัยจึงใคร่ขอขอบพระคุณมา ณ ที่นี้ด้วย

สารบัญ

บทที่
หน้า

บทที่ **หน้า**

สารบัญตาราง

ตารางที่
หน้า

สารบัญภาพ

ภาพประกอบ **หน้า**

ภาพประกอบ **หน้า**

1. บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในช่วงเวลาประมาณ 10-20 ปี ที่ผ่านมา ได้มีการคิดค้นระเบียบวิธีต่างๆมากมายเพื่อแก้ปัญหาในการรู้จำใบหน้า การรู้จำใบหน้าเป็นหนึ่งในปัญหาที่สำคัญใน การรู้จำรูปแบบ (pattern recognition) ความนิยมของการรู้จำใบหน้าเป็นผลมาจากประโยชน์ที่ได้จากสถานการณ์จริงที่เกิดขึ้นตั้งแต่ การอินเทอร์เฟสระหว่างคนกับคอมพิวเตอร์ ไปจนถึง การพิสูจน์ตน (authentication) และ การตรวจการณ์ (surveillance) ในอีกด้านหนึ่ง การรู้จำเป้าหมายอัตโนมัติ นั้นเป็นปัญหาการรู้จำรูปแบบที่มีความซับซ้อน ซึ่งนักวิจัยทางด้านนี้ให้ความสนใจค้นคว้าเป็นจำนวนมาก โดยเฉพาะทางด้านการทหาร เช่น การแบ่งคลาสยานพาหนะที่ใช้ในสงคราม การแยกชนิด เพื่อนหรือศัตรู ด้วยภาพถ่ายที่มองจากอากาศยานไร้นักบิน (Unmanned Aerial Vehicle: UAV) หรือ เรดาร์ช่องเปิดเชิงสังเคราะห์ (Synthetic Aperture Radar: SAR) หรือภาพเป้าหมายอินฟาเรด (FLIR)

ล่าสุด ระเบียบวิธี *การสร้างคืนความละเอียดสูงยวดยิ่ง* (super-resolution reconstruction: SRR) สำหรับ *ปริภูมิย่อยใบหน้า* (face subspace) ที่มีขนาดมิติเล็ก ได้ถูกเสนอสำหรับการรู้จำใบหน้า ปริภูมิย่อยใบหน้านี้ หรือที่เรียกว่า *ใบหน้าเฉพาะ* (eigenface) ถูกสกัดเป็นโดยใช้ *การวิเคราะห์ส่วนประกอบมุขหลัก* (principal component analysis: PCA) อย่างไรก็ดี หนึ่งในข้อเสียของ *ตัวบ่งต่าง* (feature) ที่สร้างคืนให้มีความละเอียดสูงยวดยิ่ง จาก ปริภูมิย่อยใบหน้า คือ ข่าวสารที่สำคัญต่อ การจำแนกคลาส (classification) ได้สูญเสียไปเนื่องจาก PCA

โดยทั่วไป คุณภาพของข้อมูล, การสกัดคุณลักษณะ (*feature extraction*), *การวิเคราะห์ความสามารถในการแบ่งแยก* (discriminant analysis) และกฎในการจัดประเภท (classification rule) เป็นสี่องค์ประกอบพื้นฐานที่สำคัญในระบบจดจำใบหน้าและเป้าหมาย หนึ่งในสิ่งที่ก่อให้สมรรถภาพระบบรู้จำดีขึ้น คือการเพิ่มคุณภาพของภาพ โดยอาจเน้นไปที่การปรับปรุงอุปกรณ์รับภาพ เพื่อให้คุณภาพของภาพ ต่อสัญญาณรบกวน ความพร่า และ ความละเอียดที่ต่ำ ดีขึ้น

ในส่วนที่เกี่ยวกับ คุณภาพของข้อมูล เมื่อความละเอียดของภาพที่ได้จากการอุปกรณ์รับภาพมีขนาดเล็กเกินไป คุณภาพของรายละเอียดหรือข่าวสารก็จะจำกัดเกินไปสำหรับการรู้จำ เป็นผลให้การตัดสินใจมีคุณภาพต่ำอย่างรุนแรงในทุกระบบของระบบการรู้จำที่มีอยู่ อย่างไรก็ดี เป็นเพราะเรามี อัลกอริทึมในการสร้างคืนความละเอียดสูงยวดยิ่ง (Park et al., 2003) ดังที่ทราบดีแล้วว่า เราสามารถสร้างคืนภาพความละเอียดสูง (High Resolution: HR) จากชุดลำดับภาพที่มี *การชักตัวอย่างน้อยเกินไป* (undersampling) โดยภาพที่มีความละเอียดต่ำเหล่านี้แต่ละภาพจะมีเหลื่อมกันของพิกเซลเกิดขึ้น เนื่องจากเซนเซอร์รับภาพจากภาพสถานที่จริง ภาพ HR นี้

สามารถนำมาใช้ป้อนเข้าระบบรู้จำเพื่อปรับปรุงประสิทธิภาพการจำแนกได้ ในความเป็นจริง ทั้งนี้ ความละเอียดสูงยวดยิ่ง สามารถถูกจัดให้อยู่ในหมวดคณิตศาสตร์ ด้านการคำนวณเชิงตัวเลข และ *เรกูลาไรเซชัน* (regularization) ของปัญหาขนาดใหญ่ที่มี *เงื่อนไขเลว* (ill-condition) ที่ใช้ในการอธิบายความสัมพันธ์ระหว่าง พิกเซลของภาพความละเอียดต่ำ (LR) และ สูง (HR) ตามลำดับได้ (Nguyen et al., 2001)

อีกด้านหนึ่งของเหรียญ การสกัดคุณลักษณะมีเป้าประสงค์ในการลดขนาดมิติของภาพใบหน้า หรือ ภาพเป้าหมาย โดยให้ คุณลักษณะที่สกัดได้มีความสามารถในการแทนรูปมากที่สุดเท่าที่จะมากได้ จะเห็นได้ว่า เป้าประสงค์ของการสกัดคุณลักษณะจะเป็นไปในทิศทางที่ต่างกันคนละขั้วกับการสร้างคืนความละเอียดสูงยวดยิ่ง ซึ่งเน้นการสร้างคืนภาพให้มีขนาดมิติเพิ่มขึ้นจากชุดลำดับภาพความละเอียดต่ำ การสร้างคืนความละเอียดสูงยวดยิ่งในระดับพิกเซลได้ถูกนำมาใช้ในการรู้จำ (Lin et al., 2005; Wagner et al., 2004) ซึ่งให้สมรรถนะในการรู้จำที่ดีขึ้น อย่างไรก็ดี ในแง่ของการปรับปรุงความซับซ้อนในการตัดสินใจ และ ความคงทนต่อการความผิดพลาดใน *การลงทะเบียนภาพ* (registration) งานวิจัยส่วนใหญ่ (Gunturk et al., 2003; Jia & Gong, 2005; Sezer et al., 2006) จึงเน้นความสนใจไปยังระเบียบวิธี การสร้างคืนความละเอียดสูงยวดยิ่งของภาพเฉพาะ

เพื่อต้องการเพิ่มสมรรถนะของระบบรู้จำในการรู้จำให้ถูกต้องมากขึ้น โดยการปรับปรุง ตัวบ่งต่าง มี *ความสามารถในการแบ่งแยก* (discriminant) มากขึ้น ในเบื้องต้น คณะผู้วิจัยมีความสนใจใน *การวิเคราะห์ส่วนประกอบหลักสองมิติ* (two-dimensional principal component analysis: 2DPCA) เนื่องจาก มีหลายงานวิจัย ที่ตีพิมพ์ผลงานแสดงว่า ความสามารถในการรู้จำใบหน้าดีขึ้นเมื่อใช้ 2DPCA ในการสกัดตัวบ่งต่าง แนวความคิดที่สำคัญของความละเอียดยวดยิ่งในระดับฟังก์ชันเฉพาะโดยการใช้ *ใบหน้าเฉพาะสองมิติ* แทนที่ *ใบหน้าเฉพาะหนึ่งมิติ* แบบดั้งเดิมนั้น เพื่อที่จะเอาชนะปัญหาที่สำคัญสามประการของระบบรู้จำใบหน้า นั้นก็คือคำสาบของมิติ การคำนวณที่ไม่เหมาะในทางปฏิบัติของการแยกย่อยค่าเอกฐานในกรณีที่ภาพใบหน้าเป็นภาพขนาดใหญ่ หรือ มีคุณภาพสูง สุดท้าย คือการทำลายโครงสร้างทางธรรมชาติ ซึ่งจะทำให้สหสัมพันธ์บางอย่างที่มีอยู่ในภาพให้เสียหายไป

ด้วยเหตุนี้ คณะผู้วิจัยสนใจระเบียบวิธีใหม่ ที่ยึดหลักการสร้างคืนความละเอียดสูงยวดยิ่งในระดับปริภูมิย่อยสำหรับวัตถุโดยการแทนที่ PCA ด้วย 2DPCA แต่พบปัญหาที่เป็นอุปสรรคหลักของการสร้างคืนความละเอียดสูงยวดยิ่งสำหรับ 2DPCA ซึ่งอยู่ที่การสร้างแบบจำลองทางคณิตศาสตร์เพื่อใช้ในการสร้างสมการเพื่อหาคำตอบ การสร้างคืนความละเอียดสูงยวดยิ่งสำหรับ 2DPCA จึงเป็นปัญหาทางคณิตศาสตร์ที่น่าสนใจ ที่ยังไม่มีนักวิจัยคนใดสนใจศึกษามาก่อน ผลการทดลองบนฐานข้อมูลใบหน้า Yale AR และ ORL ให้ผลดีเป็นที่น่าพอใจ นอกจากนี้ เรายังได้ทำการทดสอบกับฐานข้อมูลเป้าหมาย MSTAR ซึ่งเป็นยานพาหนะทางทหาร

ดังที่ ศาสตราจารย์ ทางวิศวกรรมไฟฟ้า Martin Vetterli นักวิจัยชั้นนำทางด้านผลการแปลง

เวฟเลต ได้กล่าวไว้ว่า "*ก่อนจะทำการบีบอัด เราจะต้องขยายก่อน* (To compress, we must first expand)" ส่วนงานวิจัยที่เรากำลังศึกษาในขณะนี้ เราจะใช้วลีที่สวนทางกับวลีข้างต้น คือ "*ก่อนจะทำการสร้างคืนความละเอียดสูงยวดยิ่งเพื่อการรู้จำ เราจะต้องลดขนาดของมิติก่อน*" อนึ่ง การสร้างคืนความละเอียดสูงยวดยิ่งของภาพเฉพาะ ที่นักวิจัยส่วนใหญ่สนใจกันอยู่นั้นยังอยู่ในขั้นแรกของการพัฒนา เนื่องจากระเบียบวิธีดังกล่าว เป็นการประมวลผลแบบสองขั้นตอน (two-stage method) ที่ว่าเป็น การประมวลผลแบบสองขั้นตอน เพราะว่าในขั้นแรก เราต้องทำการหาปริภูมิย่อยของ PCA ก่อน เสร็จแล้วในขั้นที่สอง เราจึงทำการสร้างคืนความละเอียดสูงยวดยิ่ง

อย่างไรก็ตาม ในอนาคตงานวิจัยควรจะดำเนินไปในทิศทางของการประมวลผลทั้งการหาปริภูมิย่อยของ PCA และการสร้างคืนความละเอียดสูงยวดยิ่งไปพร้อมกันเลย หรือ ในหัวข้อวิจัยใหม่ในชื่อว่า *การประมวลผลการทำให้ความละเอียดสูงขึ้นและการลดขนาดมิติร่วม* (joint dimensionality reduction-resolution enhancement processing)

งานวิจัยของเราได้รับแรงบันดาลใจจากงานวิจัยเกี่ยวกับ *ความเป็นไปได้ทางชีวภาพ* (bio-logical plausible) ของ การมอง (vision) มีงานวิจัยที่คล้ายคลึงกับการรู้จำใบหน้าด้วย PCA (Werblin, 2007) ที่วิจัยว่า *จอประสาทตา* (retina) ประมวลผลข่าวสารโดยการส่งวิดีทัศน์จำนวนมากไปยังสมอง ทั้งเป็นในลักษณะเชิงพื้นที่ หรือ เชิงเวลา นอกจากนี้ ในงานวิจัยที่เกี่ยวข้องยังมีการค้นพบว่า *สมรรถนะการเลือกแบบไขว้* (crossed selectivity) เกิดขึ้นหลายระดับในจอประสาทตาของกระต่าย อีกด้วย ซึ่งงานวิจัยเหล่านี้คล้องจองกับงานวิจัยนี้ ทั้งในส่วนของ การแทนรูปสัญญาณที่ได้รับจากจอประสาทตาผ่านปมประสาทไปยังส่วนประมวลผลของสมอง และ ผลรวมของการแทนรูปภาพใบหน้าที่ส่งสัญญาณไปสู่สมองในเชิงเวลา ซึ่งคณะ ผู้วิจัยรู้สึกภาคภูมิใจในหัวข้อที่ทำวิจัยนี้เป็นอย่างมาก ทั้งนี้เนื่องมาจากว่า สมองของมนุษย์เป็นระบบรู้จำที่ดีที่สุดในโลก ด้วยความคาดหวังที่ว่าหากเราสามารถทำความเข้าใจการทำงานของจอประสาทตาและสมองได้อย่างลึกซึ้ง จะเปิดโอกาสให้เราเข้าใจกระบวนการในการรู้จำได้ดีขึ้น ซึ่งจะทำให้เราสามารถสร้างระบบรู้จำที่ดีทีสุดขึ้นได้ในอนาคตอันใกล้นี้

ในบทที่ 2 เราจะกล่าวถึงทฤษฎีคณิตศาสตร์พื้นฐานที่ใช้ในงานวิจัยนี้อย่างย่อ ส่วนในบทที่ 3 รายละเอียดของระเบียบวิธีที่เรานำเสนอในการสร้างคืนปริภูมิย่อย 2DPCA ความละเอียดสูงยวดยิ่ง จะได้ถูกนำมากล่าวถึง ส่วนรายละเอียดของระเบียบวิธีในการใช้ 2DPCA รุ่นถัดมา เช่น diagonal 2DPCA และ 2DPCA แบบไขว้ ในการรู้จำใบหน้าและเป้าหมาย อัตโนมติ จะถูกนำมาแสดงในบทที่ 4 ซึ่งในบทนี้เราจะกล่าวถึง การคัดเลือก 2DPCA รุ่นถัดมา เฉพาะส่วนประกอบมุขสำคัญที่มีลักษณะพิเศษ เพื่อมาสร้างเป็น *ระบบจำแนกแบบหลายตัว* (multiple classifier system) ในบทที่ 5 ผลการทดลองบนฐานข้อมูล FERET, Yale, ORL, AR และ MSTAR จะถูกนำเสนอ ส่วน สรุปและข้อเสนอแนะ ปัญหาและงานวิจัยในอนาคต จะถูกกล่าวถึงในบทที่ 6

1.2 วัตถุประสงค์ของการวิจัย

วัตถุประสงค์หลักของโครงการวิจัยนี้ เพื่อศึกษาทั้งทางทฤษฎีและผลเชิงเลขในการปรับ ปรุงสมรรถนะของการวิเคราะห์องค์ประกอบมุขสำคัญ (Principal Component Analysis: PCA) โดยใช้การคัดเลือกตัวแปร และ การสร้างคืนความละเอียดสูงยวดยิ่งเพื่อเลือก *ส่วนประกอบมุขสำคัญสองมิติไขว้* (cross-two-dimensional principal component analysis: cross- 2DPCA) ที่ดีที่สุด โดยมีการสร้างความละเอียดสูงยวดยิ่งในระดับ *ปริภูมิลักษณะ* เฉพาะ (eigen-domain) แทนที่ที่จะทำในระดับพิกเซลซึ่งทำในกรรมวิธีดั้งเดิม ทั้งนี้เพื่อเป็นการลดการคำนวณและในขณะเดียวกันเพิ่มสมรรถนะของการรู้จำใบหน้าและเป้าหมายอัตโนมัติ

ในการศึกษานี้จะมีการศึกษา *ค่าดัชนีหรือคะแนน* (measure or score) ที่นำมาใช้ได้เพื่อเลือก cross-2DPCA ที่ดีที่สุดที่คำนวณจากเมตริกซ์ความแปรปรวนร่วมเกี่ยวไขว้สองมิติ (2D cross-covariance matrix) เมื่อนำเวกเตอร์เฉพาะ (eigen-vector) หรือ เวกเตอร์มุขสำคัญ (principal vector) ที่คัดเลือกมาคำนวณกับเมตริกซ์ภาพเพื่อสกัดลักษณ์ (feature extraction) หลังจากการสกัดลักษณ์เราจะได้ *เมตริกซ์ลักษณ์ (feature matrix)* ที่สามารถนำไปใช้เพื่อการรู้จำใบหน้าหรือเป้าหมายอัตโนมัติ ซึ่งแตกต่างจากการคำนวณ PCA แบบดั้งเดิมที่เราจะได้ *เวกเตอร์ลักษณ์* (ในการคำนวณ PCA แบบดั้งเดิมเมตริกซ์ภาพจะถูกแปลงเป็นเวกเตอร์ซึ่งเมื่อคำนวณเวกเตอร์ภาพที่ถูกแปลงมาแล้วกับเซตของเวกเตอร์เฉพาะ จะได้เวกเตอร์ลักษณ์)

ข้อได้เปรียบข้อหนึ่งของ cross-2DPCA ต่อ PCA คือ 2DPCA ที่มีความถูกต้องในการคำนวณเวกเตอร์เฉพาะมากกว่าเนื่องจากเวกเตอร์เฉพาะจะมีขนาดเล็กกว่ามาก นอกจาก นั้นคณะผู้วิจัยมีความสนใจ *นัยสำคัญส่วนประกอบมุขสำคัญสองมิติแบบไขว้* (generalized 2DPCA) คณะผู้วิจัยเป็นคณะแรกที่สามารถสร้างแบบจำลองทางคณิตศาสตร์ที่สามารถอธิบาย นัยทั่วไปของ 2DPCA ได้ครบทุกทิศทาง ซึ่งในที่นี้จะถูกเรียกว่า cross-2DPCA และจากการทดลองจะพบว่าเวกเตอร์เฉพาะบางทิศทางจะมีความถูกต้องในการรู้จำใบหน้ามากกว่าเวกเตอร์เฉพาะในทิศทางอื่น ทั้งนี้ผลการทดลองสอดคล้องกับ *ความเป็นไปได้ทางชีวภาพ* (biological plausible) ของการเห็น (vision) ซึ่งจะมีความไวต่อการเปลี่ยนแปลงทางแนวนอนมากกว่าทางแนวตั้ง จากข้อได้เปรียบของ 2DPCA แบบไขว้ ดังกล่าว

คณะผู้วิจัยกำลังขยายผลในทางทฤษฎีและการคำนวณเชิงเลข โดยรวมระเบียบวิธี *สกัดลักษณ์ความละเอียดสูงยวดยิ่งจากชุดลำดับของลักษณ์ความละเอียดต่ำ* (sequence of low-resolution features) เข้ากับผลการแปลงเวฟเลตระบุทิศทาง (directional wavelet transform)

ระเบียบวิธีสร้างคืนสำหรับการรู้จำใบหน้ามี 2 วิธี ระเบียบวิธีแรก เกี่ยวข้องกับการสร้างคืนภาพความละเอียดสูงยวดยิ่งจากชุดลำดับของภาพความละเอียดต่ำ หรือ อีกนัยหนึ่งเป็นการสร้างคืนความละเอียดสูงยวดยิ่งในระดับพิกเซล หลังจากนั้นนำชุดเรียนรู้ (training set) ของภาพความละเอียดสูงยวดยิ่งที่สร้างคืนเรียบร้อย นำมาคำนวณหาค่าเวกเตอร์เฉพาะ

ส่วนในระเบียบวิธีที่สอง เวกเตอร์เฉพาะความละเอียดสูงยวดยิ่งจะถูกสร้างคืนโดยตรงจากชุดเวกเตอร์เฉพาะที่คำนวณจากชุดเรียนรู้ (training set) ของภาพความละเอียดต่ำที่มีตัวกระทำการแปลงภาพที่เหมือนกัน หรือ อีกนัยหนึ่งคือการสร้างคืนเวกเตอร์เฉพาะความละเอียดสูงยวดยิ่งในระดับ *ปริภูมิเฉพาะ* (eigenspace-domain) ซึ่งในที่นี้ *เมตริกซ์ความแปรปรวนร่วมเกี่ยวไขว้สองมิติ* (2D cross-covariance matrix) จะถูกนำมาใช้คำนวณร่วมกับสมการความละเอียดสูงยวดยิ่งตามระเบียบวิธีที่ศึกษาและพัฒนาขึ้นนี้ เราจะได้ *N* ชุดของ เวกเตอร์เฉพาะความละเอียดสูงยวดยิ่งที่คำนวณจากจำนวน *N* แบบ ของ 2D cross-covariance matrix ทั้งหมด เนื่องจาก เวกเตอร์เฉพาะแต่ละเวกเตอร์จาก *N* เวกเตอร์ จะมีความถูกต้องในการรู้จำที่แตกต่างกัน หรืออีกนัยหนึ่งเวกเตอร์เฉพาะบางตัว มีความสามารถในการสกัดลักษณ์ที่เหมาะกับการรู้จำใบหน้าได้ดีกว่าเวกเตอร์เฉพาะที่เหลือ (ผลจากตัวบ่งต่างที่อาจอ่อนไหวในทิศทางหนึ่งมากกว่าทิศทางอื่น) ด้วยเหตุนี้เราสามารถสร้างระบบรู้จำใบหน้าและเป้าหมายอัตโนมัติ จากเวกเตอร์เฉพาะที่สกัดได้ เป็น 2 วิธี

วิธีแรก โดยการหาเวกเตอร์เฉพาะที่ดีที่สุดเพียงเวกเตอร์เดียวเพื่อสร้าง *ตัวจำแนกแข็ง แกร่ง* (strong classifier) สำหรับรู้จำใบหน้า หรือ วิธีที่สองใช้เวกเตอร์เฉพาะมากกว่าหนึ่งตัวเพื่อสร้าง *ตัวจำแนกอ่อน* (weak classifier) จำนวนหนึ่ง และ รวมผลของการตัดสินเพื่อสร้าง *ระบบตัวจำแนกแบบหลายตัว* (multiple classifier systems) ซึ่งที่กล่าวถึงข้างต้นทั้งหมดเป็นครึ่งหนึ่งของงานวิจัยของโครงการนี้

สำหรับอีกครึ่งหนึ่งของโครงการวิจัยเป็นการศึกษา *การไขว้เฉพาะที่เหมาะสม* (cross-selectivity) ที่คำนวณโดยตรงจาก 1D cross-covariance matrix เพื่อให้ได้ 2D cross-covariance matrix ใหม่ขึ้น โดยคำนึงถึงสมาชิกพื้นฐานเพื่อนบ้าน (neighborhood element) ให้มากขึ้นโดยการใช้ผลการแปลงเวฟเลตระบุทิศทาง (directional wavelet) จะทำให้ได้ 2DPCA แบบไขว้ในขณะเดียวกันกับที่มีการวิเคราะห์แบบหลายระดับความละเอียด (ราย ละเอียดในส่วนนี้จะกล่าวถึงต่อไปในหัวข้อย่อยต่อไป) อนึ่งเราสามารถทำการสร้างคืนเวกเตอร์เฉพาะความละเอียดสูงยวดยิ่งของ cross-2DPCA ชนิดใหม่นี้ได้ตามระเบียบวิธีที่กล่าวมาก่อนนี้ได้เช่นกัน

1.3 ขอบเขตของการวิจัย

1.การวิเคราะห์ปริภูมิย่อย 2DPCA สุ่ม
2.ทดสอบการรู้จำใบหน้าและเป้าหมายอัตโนมัติ ด้วยการคัดเลือกตัวแปรที่ดีที่สุด ของ 2DPCA ไขว้
3.สร้างแบบจำลองทางคณิตศาสตร์ของการสร้างคืนลักษณ์ความละเอียดสูงยวดยิ่ง จากชุดลำดับของลักษณ์ความละเอียดต่ำด้วย 2DPCA ไขว้ สำหรับการรู้จำใบหน้าและเป้าหมายอัตโนมัติ

4.ขยายองค์ความรู้ด้านผลการแปลงเวฟเลตระบุทิศทาง เพื่อสร้าง 2DPCA ไขว้ เพื่อขยายความให้มีนัยทั่วไป

5.ศึกษาความเป็นไปได้ของ การเลือกถ่วงน้ำหนักเล็งเลิศสำหรับ 2DPCA ไขว้ และ 2DPCA ไขว้แบบหลายระดับความละเอียด

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. ฐานข้อมูลใบหน้าและเป้าหมายสำหรับใช้ทดสอบระเบียบวิธีขั้นสูงในอนาคต
2. โปรแกรมรู้จำใบหน้าและเป้าหมายอัตโนมัติโดยระเบียบวิธีปริภูมิย่อย 2DPCA แบบไขว้
3. องค์ความรู้ด้านการวิเคราะห์ส่วนประกอบมุขสำคัญสองมิติไขว้ และระบบจำแนกหลายตัวโดยการถ่วงน้ำหนักอย่างเล็งเลิศ
4. ระเบียบวิธี การวิเคราะห์ปริภูมิย่อย 2DPCA สุ่ม (random 2DPCA ปริภูมิย่อย analysis)
5. องค์ความรู้ และ แบบจำลองทางคณิตศาสตร์ ของการสร้างคืนลักษณ์ความละเอียดสูงยวดยิ่งจากชุดลำดับของลักษณ์ความละเอียดต่ำด้วย 2DPCA ไขว้
6. องค์ความรู้ด้านผลการแปลงเวฟเลตระบุทิศทาง
7. บทความวิชาการเสนอในที่ประชุมนานาติ และ วารสารวิชาการในระดับนานาชาติ

1.5 ลำดับขั้นตอนในการวิจัย

1. การจัดเตรียมฐานข้อมูลใบหน้าและเป้าหมาย (YALE AR ORL และ MSTAR)
2. นอกจากนี้ยังได้มีการเตรียมฐานข้อมูลใบหน้าขนาดใหญ่ (FERET) ที่เพิ่งได้รับอนุญาตจาก NIST
3. ศึกษาลักษณะเฉพาะทิศทางของการวิเคราะห์ส่วนประกอบมุขสำคัญสองมิติไขว้
4. ศึกษาและทดสอบ เกณฑ์ที่ใช้ในการคัดเลือก ชุดของสัมประสิทธิ์ของ 2DPCA ไขว้
5. ศึกษาและทดสอบการรู้จำใบหน้าและเป้าหมายอัตโนมัติ ด้วยการปรับปรุงความละเอียดสูงยิ่งยวดในระดับปริภูมิเฉพาะ (eigenspace-domain) ของ 2DPCA ไขว้
6. สรุป เขียน และ นำเสนอ บทความในที่ประชุมวิชาการในระดับนานาชาติ
7. ศึกษาผลการแปลงเวฟเลตระบุทิศทาง เพื่อนำมาร่วมใช้กับ 2DPCA
8. ศึกษาและทดสอบการรู้จำใบหน้าและเป้าหมายอัตโนมัติ ด้วยการสร้างคืนปริภูมิเฉพาะความละเอียดสูงยิ่งยวด 2DPCA ด้วยผลการแปลงเวฟเลตแบบระบุทิศทาง
9. สรุป เขียน และนำเสนอ บทความในที่ประชุมวิชาการ หรือ วารสารวิชาการ ในระดับนานาชาติ

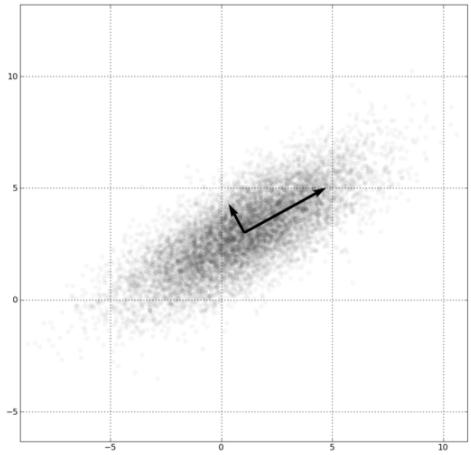
2. ทฤษฎีคณิตศาสตร์พื้นฐาน

บทนี้ บรรยายเกี่ยวกับ ทฤษฎีคณิตศาสตร์พื้นฐานที่เกี่ยวข้อง กับ การรู้จำใบหน้าและเป้าหมายอัตโนมัติ โดยจะกล่าวถึงระเบียบวิธี การสกัดตัวบ่งต่าง ด้วย การวิเคราะห์ส่วนประกอบมุขสำคัญ (PCA) เพื่อใช้กับระบบรู้จำ รวมถึงการปรับปรุง ปัญหา และอุปสรรคของ PCA เพื่อนำไปสู่ PCA รุ่นถัดมา เช่น 2DPCA และ 2DPCA แบบไขว้ เป็นต้น นอกจากนั้น เหตุผลและแรงบันดาลใจในการชี้ชัดว่า PCA เป็นตัวบ่งต่างที่เหมาะสม จะถูกขยายให้เห็นชัดเจนขึ้นผ่านทางความสัมพันธ์ระหว่าง PCA กับ ความเป็นไปได้ทางชีวภาพของการมอง จากนั้น เราจะกล่าวถึงระเบียบวิธี การสร้างคืนความละเอียดสูงยวดยิ่ง เมื่อเราได้ศึกษาถึงรายละเอียดของระเบียบวิธีทั้งสองดีแล้ว เราก็จะกล่าวถึง การสกัดตัวบ่งต่างความละเอียดสูงยวดยิ่งในระดับของปริภูมิย่อย ที่เหมาะสมมากขึ้นสำหรับการรู้จำใบหน้าและเป้าหมายอัตโนมัติ

และเพราะเราจะนำ 2DPCA แบบไขว้ มาใช้เป็นตัวบ่งต่าง ซึ่งเราได้พบว่า ตัวบ่งต่างที่สกัดได้ เมื่อนำมาใช้กับระบบรู้จำ ระบบจะตอบสนองต่อทิศทางของ 2DPCA ที่ใช้ เราจึงต้องใช้การคัดเลือกตัวบ่งต่าง โดยคาดหวังว่าเมื่อเราเลือกเฉพาะส่วนหนึ่งของ 2DPCA ไขว้แล้ว ระบบรู้จำจะมีความถูกต้องในการรู้จำมากขึ้น เหตุผลในการที่เราใช้ชุดของ 2DPCA แบบไขว้ เป็นเพราะสมมติฐานตามความเป็นไปได้ทางชีวภาพที่ว่า ปมประสาทที่เกี่ยวข้องกับการรู้จำนั้นมีความซ้ำซ้อนของข่าวสารกันอยู่ และ ช่วยกันในการรู้จำ สุดท้ายเราจะกล่าวถึง ผลการแปลงเวฟเลตแบบระบุทิศทาง ซึ่งมีวัตถุประสงค์ที่จะนำใช้กับ 2DPCA เพื่อปรับปรุงสมรรถนะในการรู้จำ

2.1 การวิเคราะห์ส่วนประกอบมุขสำคัญ

การวิเคราะห์ส่วนประกอบมุขสำคัญ (Principal Component Analysis: PCA) คือ กระบวนการทางคณิตศาสตร์ที่ใช้ *การแปลงเชิงตั้งฉาก* (orthogonal transformation) ในการเปลี่ยนชุดของข้อมูลสังเกตการณ์ที่อาจเป็นตัวแปรสุ่มที่มี *สหสัมพันธ์* (correlation) ต่อกัน ให้เปลี่ยนเป็นชุดของตัวแปรสุ่มที่ไม่สหสัมพันธ์ต่อกัน ซึ่งถูกเรียกว่า *ส่วนประกอบมุขสำคัญ* (principal component) ในทางปฏิบัติ จำนวนของส่วนประกอบมุขสำคัญจะมี *มิติ* (dimension) ที่น้อยกว่าหรือเท่ากับมิติดั้งเดิมของตัวแปรต้นฉบับ การแปลงนี้ถูกนิยามในลักษณะที่ส่วน ประกอบมุขสำคัญแรกจะมีค่าความแปรปรวนสูงที่สุด (นั้นคือ เป็นส่วนประกอบสำคัญที่มีการแปรผันของข้อมูลมากที่สุด) และ ส่วนประกอบถัดไปจะมีความแปรปรวนของข้อมูลที่สูงสุดถัดไปทั้งนี้เป็นไปภายใต้ข้อจำกัดที่ว่า **ส่วนประกอบนั้นจะต้องตั้งฉากกับส่วนประกอบก่อนหน้านี้ หรือ อีกนัยหนึ่ง คือไม่มีส่วนประกอบจะไม่มีสหสัมพันธ์ต่อกัน** การที่ส่วนประกอบมุขสำคัญจะเป็นอิสระต่อกันอย่างสมบรูณ์จะเกิดขึ้น เฉพาะในกรณีที่ชุดข้อมูลมีการแจกแจงเป็น



รูปที่ 2.1 ส่วนประกอบมุขสำคัญ ที่ตั้งฉากซึ่งกันและกัน

การแจกแจงปกติร่วม (joint normal distribution) ข้อด้อยของ PCA คือ PCA นั้นมีความอ่อนไหวต่อ *การสเกลสัมพัทธ์* (relative scaling) ของตัวแปรต้นฉบับ PCA มีชื่อกันต่างกันออกไปขึ้นอยู่กับการประยุกต์ใช้งาน เช่น *ผลการแปลง* Karhunen–Loève (KLT), *ผลการแปลง* Hotelling หรือ *การแยกย่อยเชิงตั้งฉากที่เหมาะสม* (proper orthogonal decomposition: POD)

PCA ถูกสร้างขึ้นในปี ค.ศ. 1901 โดย Karl Pearson ขณะนี้ PCA ส่วนใหญ่จะถูกใช้เป็นเครื่องมือใน *การวิเคราะห์ข้อมูลเชิงวินิจฉัย* (exploratory data analysis) และ สำหรับการทำ *แบบจำลองทำนาย* (predictive model) PCA สามารถคำนวณ โดย *การแยกย่อยค่าลักษณะเฉพาะ* (eigenvalue decomposition) ของเมทริกซ์ค่าความแปรปรวนร่วมเกี่ยวของข้อมูล หรือ *การแยกย่อยค่าเอกฐาน* (singular value decomposition: SVD) ของเมทริกซ์ข้อมูล ซึ่งโดยปกติจะกระทำหลังจากทำการปรับข้อมูลให้มีค่าเฉลี่ยเท่าศูนย์ในแต่ละมิติแล้ว

PCA มักถูกพิจารณาว่าเป็นเครื่องมือที่สามารถเปิดเผยคุณสมบัติภายในของโครงสร้างของข้อมูล โดยเฉพาะในแง่ที่สามารถอธิบายความแปรปรวนของข้อมูลได้ดีที่สุด ในกรณีที่ ข้อมูลแบบหลายตัวแปร ข้อมูลจะถูกมองว่าเป็นจุดที่อยู่ใน *ชุดพิกัด* (set of coordinates) ของปริภูมิข้อมูลที่มี *มิติขนาดใหญ่มาก* (high dimensionality) ในที่นี้ 1 แกนพิกัดจะเท่ากับหนึ่งตัวแปร PCA สามารถให้เอาท์พุตในรูปแบบที่มี มิติขนาดเล็ก (low dimensionality) ลง จากรูปที่ 2.1 "เงา" ที่ปรากฏนั้น ให้ความหมายที่มีประโยชน์อย่างมาก ดังจะเห็นได้ว่า โดยการใช้เพียงส่วนประกอบหลักสองสามส่วนแรก (สองสามแกนพิกัดใหม่) จะได้ว่า ขนาดมิติของข้อมูลที่ถูกแปลงจะมี ขนาดมิติที่ลดลง

มุมมองของ PCA ต่อไปนี้ เป็นการตีความหมายของ PCA เป็นกรณีพิเศษ โดยพิจารณาข้อมูลเป็น *ตัวแปรสุ่ม* (random variable) ตามทฤษฎีของ *กระบวนการสโทรแคสติก* (stochastic process) แล้ว ทฤษฎีบท Karhunen–Loève (ซึ่งได้ถูกตั้งชื่อเพื่อเป็นเกียรติต่อ

Kari Karhunen และ Michel Loève) คือการแทนรูปกระบวนการสโทรแคสติกด้วยผลรวมเชิงเส้นของ *ฟังก์ชันเชิงตั้งฉาก* (orthogonal) จำนวนอนันต์ อุปมานเช่นเดียวกับการแทนรูปฟังก์ชันช่วงจำกัดด้วย *อนุกรมฟูริเยร์* (Fourier Series) อนึ่งกระบวน การสโทรแคสติก ที่ปรากฏในรูปของอนุกรมลำดับอนันต์รูปแบบนี้ได้มีการศึกษามาก่อนหน้านี้โดย Dharmananda Damodar Kosambi *การกระจาย* (expansion) กระบวนการสโทรแคสติกมีหลากหลายวิธี หากกระบวนการถูกกำหนดให้อยู่ในช่วง $[a,b]$ แล้ว ฟังก์ชันฐานหลักเชิงตั้งฉากใดก็ตามที่สามารถแผ่ได้ทั่วปริภูมิ $L^2([a,b])$ ก็สามารถแทนรูปกระบวนการดังที่กล่าวมาแล้วได้ ความ สำคัญของทฤษฎีบท Karhunen–Loève อยู่ที่ว่า เราสามารถหาฟังก์ชันฐานหลักที่ดีที่สุดในความหมายที่ว่าฟังก์ชันฐานหลักที่หาได้จะให้ค่า *ความผิดพลาดเฉลี่ยกำลังสอง* (mean square error) ที่มีค่าน้อยที่สุดได้

อนุกรมฟูริเยร์มีสัมประสิทธิ์ของการกระจายเป็นจำนวนจริง และฟังก์ชันฐานหลักของการกระจายประกอบด้วย ฟังก์ชันไซน์หลากหลายความถี่จำนวนหนึ่ง (ที่จริงประกอบด้วยทั้งฟังก์ชันไซน์และโคไซน์) ในทางตรงกันข้าม **ตามทฤษฎีบท Karhunen–Loève สัมประสิทธิ์ที่ได้จะเป็น *ตัวแปรสุ่ม* (random variables) และ มีฟังก์ชันฐานหลักของการกระจายที่ปรับ ตัวตามกระบวนการ** ที่จริง ฟังก์ชันฐานหลักเชิงตั้งฉากที่ใช้ในการแทนรูปภาพนี้จะถูกกำหนดโดย ฟังก์ชันความแปรปรวนร่วมเกี่ยว (covariance function) ของกระบวนการ ในอีกแง่หนึ่งผลการแปลง Karhunen–Loève มีการปรับตัวตามกระบวนการ เพื่อให้เกิดฟังก์ชันฐานหลักที่ดีที่สุดสำหรับการกระจายกระบวนการ

ในกรณีที่กระบวนสโทรแคสติกมีศูนย์กลางอยู่ที่พิกัดจุดกำเนิด $\{X_t\}_{t\in[a,b]}$ ในที่นี้ศูนย์กลางที่พิกัดจุดกำเนิดของกระบวนการ หมายถึง *ค่าความคาดหวัง* (expectation) $E(X_t)$ มีค่าเท่ากับศูนย์ตลอดช่วงพารามิเตอร์ของ t ใน $[a,b]$ ดังนั้น X_t จะมีการแยกย่อยดังนี้

$$X_t = \sum_{k=1}^{\infty} Z_k e_k(t)$$

โดยที่ z_k เป็นตัวแปรสุ่มที่ไม่มีสหสัมพันธ์ต่อกัน และ e_k เป็นฟังก์ชันต่อเนื่องที่มีค่าเป็นจริงในช่วง $[a,b]$ และตั้งฉากซึ่งกันและกันใน $L^2([a,b])$ เป็นที่น่าสังเกตว่า การกระจายนี้ เป็นการกระจายเชิงตั้งฉากเชิงทวิ คือ z_k ตั้งฉากกันใน **ปริภูมิความน่าจะเป็น** (probability space) ส่วนฟังก์ชัน e_k ตั้งฉากกันใน **ปริภูมิเวลา** กรณีทั่วไปของกระบวนการ X_t ที่ศูนย์กลางของกระบวนการไม่อยู่ที่พิกัดจุดกำเนิด สามารถแปลงกลับมาเป็นกระบวนสโทรแคสติกที่มีศูนย์กลางอยู่ที่พิกัดจุดกำเนิดได้โดยการคำนวณ $X_t - E(X_t)$ อนึ่ง ถ้ากระบวนการเป็นแบบเกาส์แล้ว ตัวแปรสุ่ม z_k ก็จะมีการกระจายแบบเกาส์ และเป็นกระบวนการสโทรแคสติกที่เป็นอิสระต่อกัน PCA ได้ถูกนำมาใช้ในการสกัดตัวบ่งต่างสำหรับการรู้จำใบหน้า ซึ่งโดยทั่วไป ค่าสัมประสิทธ์ของ PCA ของภาพใบหน้า นิยมถูกตั้งสมมติฐานให้เป็นตัวแปรสุ่มที่มีการกระจายตัวแบบเกาส์ อย่างไรก็ดี

โดยนัยทั่วไป ค่าสัมประสิทธ์ของ PCA ของภาพใบหน้าสามารถกระจายตัวแบบอื่นๆ ที่มีค่าทางสถิติอันดับสูงๆ ประกอบอยู่ด้วย คณะผู้วิจัยได้สังเกตเห็นว่า ในขณะที่ ค่าสัมประสิทธิ์ของ PCA อันดับแรกๆ จะมีการกระจายตัวเป็นแบบเกาส์ ในขณะที่ ค่าสัมประสิทธิ์ของ *การวิเคราะห์ส่วนประกอบไม่สำคัญ* (minor component analysis) น่าจะเป็นตัวแปรสุ่มที่มีการกระจายตัวเป็นแบบลาปราส หรือ อาจมีการกระจายตัวเป็นแบบอื่น ที่มีค่าทางสถิติอันดับสาม และ สี่ ไม่เท่ากับศูนย์ โดยการตั้งสมมติฐานให้มีนัยทั่วไป ในอนาคตเราจะสามารถคิดงานวิจัยในใหม่ที่เกี่ยวกับการสร้างคืนความละเอียดสูงยวดยิ่งในระดับปริภูมิย่อย PCA

นอกจากนี้ คำตอบที่ผ่อนปรนของ *การจัดกลุ่มด้วย K-ค่าเฉลี่ย* (K-mean clustering) (PCA, n.d.) สามารถหาได้ด้วยจากส่วนประกอบมุขสำคัญ PCA และ ปริภูมิย่อย PCA ที่แผ่ตามทิศทางมุขสำคัญจะมีเอกลักษณ์ เท่ากับ ปริภูมิย่อยของเซนทอยด์กลุ่มที่ถูกกำหนดด้วย *เมทริกซ์ระหว่างคลาส* (between-class matrix) PCA จะฉายปริภูมิย่อย ไปตามแนว *คำตอบวงกว้าง* (global solution) ของ การจัดกลุ่มด้วย K-ค่าเฉลี่ย ด้วยเหตุนี้ PCA จะช่วยให้เราสามารถหาคำตอบใกล้เล็งเลิศ (near optimum) ของ การจัดกลุ่มด้วย K-ค่าเฉลี่ย

การกระจายข้างต้นที่ทำให้เกิดตัวแปรสุ่มที่ไม่มีสหสัมพันธ์กันนี้ เป็นที่รู้จักกันดีว่า คือการกระจาย Karhunen–Loève หรือ การแยกย่อย Karhunen–Loève สำหรับรูปแบบเชิงประจักษ์ (empirical version) ของการวิเคราะห์ส่วนประกอบของกระบวนการสโทรแคสติกนี้ สามารถคำนวณได้จากตัวอย่าง หรือ ตัวอย่างประชากร ซึ่งเป็นที่รู้จักกันดีในชื่อว่า คือผลการแปลง Karhunen–Loève (KLT) การวิเคราะห์ส่วนประกอบมุขสำคัญ การแยกย่อยเชิงตั้งฉากที่เหมาะสม (proper orthogonal decomposition) ฟังก์ชันตั้งฉากเชิงประจักษ์ (คำที่นิยมใช้ในอุตุนิยมวิทยาและธรณีฟิสิกส์) หรือ ผลการแปลง Hotelling

2.1.1 มิติเดียว

PCA แทนรูปภาพใบหน้าด้วยผลเฉลี่ยที่มีการถ่วงน้ำหนัก ใบหน้าลักษณ์ เราจะทำการแทนรูปชุดภาพใบหน้า ด้วยเมทริกซ์ขนาด $N \times M$ เมทริกซ์ ประกอบด้วยภาพใบหน้า x_i ที่ถูกทำให้เป็นเวกเตอร์ $\vec{l}_i$ และ วางต่อกันในแถวตั้งต่อเนื่องกันไป หรือ $L = \left[\vec{l}_1, \cdots, \vec{l}_M\right]$ โดยที่ N เป็นจำนวนพิกเซลทั้งหมดของภาพใบหน้า และ M เป็นจำนวนตัวอย่างที่ใช้ในการฝึกฝน ภาพใบหน้าเฉลี่ยสามารถคำนวณได้ ดังนี้

$$\overline{m_l} = \frac{1}{\mathrm{M}} \sum_{i=1}^{M} l_i \qquad (1)$$

โดยการขจัดค่าเฉลี่ยของใบหน้าออกจากภาพใบหน้าแต่ละหน้า เราจะได้

$$L=\left[\overline{l_1}-\overline{m_l},\cdots,\overline{l_M}-\overline{m_l}\right]=\left[\overline{l_1}',\cdots,\overline{l_M}'\right] \quad (2)$$

ชุดของเวกเตอร์เฉพาะนี้ จะถูกเรียกว่า **ใบหน้าลักษณ์** ใบหน้าลักษณ์นี้ คำนวณจาก *คณะ* หรือ เอนเซมเบล (ensemble) ของเมทริกซ์ความแปรปรวนร่วมเกี่ยว

$$C=\sum_{i=1}^{M}\left(\overline{l_i}-\overline{m_l}\right)\left(\overline{l_i}-\overline{m_l}\right)^T \quad (3)$$

การคำนวณหา เวกเตอร์เฉพาะ จากเมทริกซ์ C ตรงๆ หรือ ที่เรียกว่า การคำนวณจาก *ผลคูณภายนอก* (outer product) ไม่เหมาะสมในทางปฏิบัติ ทั้งนี้เนื่องมาจากว่า ขนาดของเมทริกซ์ C จะใหญ่มาก ในการแยกส่วนประกอบของ PCA เมทริกซ์ C จะถูกแยกย่อยด้วย การแยกย่อยค่าเอกฐาน ซึ่งจำเป็นต้องใช้การคำนวณและขนาดหน่วยความจำที่ใหญ่มาก หากเมทริกซ์มีขนาดใหญ่ การคำนวณจะซับซ้อนเกินกว่าที่จะคำนวณได้ ด้วยเหตุนี้ จึงได้มีการนำเสนอ (Fukunnaka, 1991; Turk, 1991) การคำนวณผลประมาณ เวกเตอร์เฉพาะ ด้วย *ผลคูณภายใน* (inner product) ของเมทริกซ์ $R=L^TL$ ที่มีขนาดเล็กกว่ามาก

$$(L^TL)V_l = V_l\Lambda_l \quad (4)$$

โดยที่ V_l คือ เมทริกซ์ลักษณ์ และ Λ_l คือ เมทริกซ์ที่มี ค่าเฉพาะ (eigenvalue) โดยการคูณทั้งสองข้างของสมการด้วย L เราจะได้

$$(LL^T)LV_l = LV_l\Lambda_l \quad (5)$$

ดังนั้น เมทริกซ์เชิงตั้งฉาก ของเวกเตอร์เฉพาะ ของ $C = LL^T$ สามารถคำนวณได้ดังนี้

$$\mathbf{\Phi}=LV_l\Lambda_l^{-\frac{1}{2}} \quad (6)$$

สำหรับภาพใบหน้าแต่ละภาพ x_l เวกเตอร์น้ำหนัก $\mathbf{a}_l$ สามารถคำนวณได้จาก *การฉาย* (projecting) ภาพใบหน้าที่ถูกแปลงเป็นเวกเตอร์ ลงบนใบหน้าลักษณ์ จำนวน *K* ใบหน้า $\mathbf{\Phi}=\left[e_1,\cdots,e_K\right]$

$$\mathbf{a}_l=\mathbf{\Phi}\left(\vec{l}_l-\overline{m_l}\right) \quad (7)$$

สมการที่ (7) คือ การแทนรูปภาพใบหน้าด้วย ใบหน้าลักษณ์ ภาพใบหน้าสามารถสร้างคืนได้จาก ใบหน้าลักษณ์

$$r_l=\mathbf{\Phi}\mathbf{a}_l+\overline{m_l} \quad (8)$$

e_1 e_2 e_3 e_4 e_5 e_{10} e_{50} e_{100} e_{200} e_{300} e_{400} e_{500} e_{600}

รูปที่ 2.2 ตัวอย่างใบหน้าลักษณ์ (eigenface) ที่เรียงตามขนาดของ ค่าเฉพาะ (eigenvalue)

เช่นเดียวกับระเบียบวิเคราะห์ *แบบหลายสเกล* (multiscale) PCA แยกย่อยภาพใบหน้าออก เป็นหลายแถบความถี่ ดังแสดงในรูปที่ 2.2 ใบหน้าลักษณ์ ที่มี ค่าเฉพาะสูง จะมีความเหมือนกับภาพใบหน้าเดิมและมีลักษณะสมบัติของส่วนประกอบความถี่ต่ำ ส่วน ใบหน้าลักษณ์ที่มีค่าเฉพาะต่ำ สังเกตแล้วจะเหมือนสัญญาณรบกวน ซึ่งมีลักษณะสมบัติของส่วนประกอบความถี่สูง เนื่องจาก PCA เป็นการแทนรูปภาพใบหน้าที่ดีที่สุด ใบหน้าลักษณ์ *K* ใบหน้าที่มีค่าเฉพาะมากที่สุด เป็นส่วนที่กำลังงานมากที่สุด จึงคงไว้ซึ่งข่าวสารที่สำคัญที่สุดของชุดข้อมูลใบหน้าภาพ ส่วนใบหน้าลักษณ์ อันดับที่มากกว่า *K* จะถูกตั้งสมมติฐานว่าเป็นสัญญาณรบกวน ความเปลี่ยนแปลงของแสงเงา และ อื่นๆ ที่นอกเหนือจากโครงสร้างใบหน้า เช่นเดียวกับการประมวลผลภาพแบบหลายสเกล ที่กำลังงานส่วนใหญ่ของภาพจะอยู่ที่แถบความถี่ต่ำ ส่วนแถบความถี่สูงจะแทนรูปขอบของภาพ และ แถบความถี่ที่สูงมากขึ้นไปจะแทนรูปสัญญาณรบกวน การที่ PCA สามารถดำรงไว้ซึ่งคุณสมบัติของแถบความถี่สูงและต่ำของภาพใบหน้าได้นั้น น่าจะมาจากเหตุผลที่ว่า เมทริกซ์ความแปรปรวนร่วมเกี่ยว เมื่อถูกแปลงฟูเรียร์แล้วจะให้ค่าสเปกตรัมทางความถี่ ดังนั้น จะเห็นได้ว่า PCA สามารถแบ่งแยกส่วนประกอบทางความถี่โดยการประมวลผลในเชิงเวลาและต้องใช้ข้อมูลที่ใช้ในการฝึกฝนในการคำนวณ ในขณะที่การวิเคราะห์แบบหลายสเกล นิยมใช้ตัวกรองทางความถี่ หรือ ฟังก์ชันฐานหลักที่มีผลตอบสนองทางความถี่ ต่างๆกัน ในการประมวลผล โดยที่ฟังก์ชันฐานหลัก ได้มีการกำหนดขึ้นมาล่วงหน้า ซึ่งไม่มีส่วนเกี่ยวข้องกับจำนวนข้อมูลที่ใช้ฝึกฝน

2.1.2 สองมิติ

เร็วๆนี้ Yang (Yang et al., 2004) เป็นนักวิจัยทีมแรกที่ได้เสนอเทคนิค ที่เรียกว่าการ *วิเคราะห์ส่วนประกอบมุขสำคัญสองมิติ* (two-dimensional principal component analysis: 2DPCA) ที่ซึ่ง เมทริกซ์ความแปรปรวนร่วมเกี่ยวของภาพใบหน้า จะถูกคำนวณโดยตรงจาก เมทริกซ์ของภาพ เพื่อที่จะได้คงไว้ซึ่งข่าวสารทั้งในด้านปริภูมิตำแหน่งและโครงสร้างของภาพใบหน้า

Yang และคณะ (Yang et al., 2004) ได้ทำการทดลองและพบว่าสมรรถนะของ 2DPCA ดีกว่า PCA ในการทดลองกับฐานข้อมูลใบหน้าหลายฐานข้อมูล

หนึ่งในความได้เปรียบของระเบียบวิธีนี้อยู่ที่ว่า ขนาดมิติของเมทริกซ์ความแปรปรวนร่วมเกี่ยว ใหม่นี้ อาจมีขนาดมิติเท่ากับ ความกว้างของภาพใบหน้า หรือ ความสูงของภาพใบหน้า (ในกรณี ของระเบียบวิธีรุ่นที่พัฒนาต่อจาก 2DPCA) ซึ่งจะมีผลดีในการหาค่า เวกเตอร์เฉพาะที่จะสามารถคำนวณได้รวดเร็วกว่ามาก

เป็นที่ควรจะจำไว้เป็นกรณีพิเศษว่า ขนาดมิติของ เมทริกซ์ความแปรปรวนร่วมเกี่ยวนี้ จะมีขนาดเล็กกว่า เมทริกซ์ที่ต้องใช้ในการคำนวณ PCA ตามปกติ ที่คำนวณด้วยผลคูณภายนอก หากพิจาณาจากทฤษฎีความน่าจะเป็นแล้ว เมทริกซ์ความแปรปรวนร่วมเกี่ยว ที่คำนวณจากภาพใบหน้า หรือ เรียกให้ถูกต้องว่า **เมทริกซ์ความแปรปรวนร่วมเกี่ยวเชิงประจักษ์ (empi-rical covariance matrix)** นี้ เป็นการประมาณ **เมทริกซ์ความแปรปรวนร่วมเกี่ยวที่แท้จริง (true covariance matrix)** ของปริภูมิภาพใบหน้า หากเราจะประมาณเมทริกซ์จริงนี้ให้มีความถูกต้อง เราจำเป็นจะต้องใช้ภาพใบหน้าจำนวนมากในการคำนวณ ซึ่งเป็นไปได้ยากในทางปฏิบัติ

ในขณะที่เมทริกซ์ความแปรปรวนร่วมเกี่ยวของ 2DPCA นั้นจะมีขนาดเล็กกว่ามาก ทำให้การประมาณ (estimation) เมทริกซ์ได้เต็ม rank หรือ มีค่าใกล้เคียงความเป็นจริงดีกว่า โดยเฉพาะใช้ชุดข้อมูลฝึกฝนที่น้อยกว่ามาก ซึ่งเหมาะสมกับการรู้จำใบหน้าเป็นอย่างมาก ฉะนั้น เมื่อเราสามารถประมาณได้ใกล้เคียงกับความเป็นจริงได้มากกว่า จะทำให้เราสามารถหลีกเลี่ยงปัญหาสำคัญในการรู้จำ ซึ่งก็คือปัญหา *จำนวนฝึกฝนน้อย* (small sample size: SSS)

นอกจากเหตุผลเกี่ยว SSS ที่กล่าวมาข้างต้นแล้ว สมรรถนะที่ดีขึ้นของ 2DPCA สามารถอธิบายด้วยหลักการที่น่าเชื่อถือกว่าได้อีกหลายหลักการ หนึ่งในหลักการนี้ ได้แก่ *คมมีดของออกแคม* (Occam's razor) ซึ่งตั้งไว้เป็นเกียรติแก่ William of Ockham (Occam, n.d.) ที่มีชีวิตอยู่ในช่วงปี ค.ศ. 1285-1349 หลักของคมมีดของออกแคมสรุปได้คร่าวๆ ดังนี้ "สมมติฐาน (hypothesis) หรือ คำอธิบายต่อปรากฏการณ์ที่พื้นฐานที่สุด คือ คำตอบที่มีน่าจะถูกต้องที่สุด" ซึ่งแปลจากภาษาอังกฤษโดยตรง ดังนี้ "The simplest explanation is most likely the correct one" หรือ คำนิยามที่นิยมใช้กันมากที่ว่า "Simpler explanations are, other things being equal, generally better than more complex ones" หรือ คำกล่าวที่ใกล้เคียงกับแนวความคิดดั้งเดิมที่ว่า "Plurality must never be posited without necessity" โดยที่ ความจำเป็นที่แท้จริง (truly necessary entity) ในที่นี้ หมายถึง *พระเจ้า* (God) ถ้าใครเคยดูภาพยนต์เรื่อง Contact ซึ่งแสดงโดย Jodi Foster จะรู้ว่า Occam's razor เป็นแก่นสารของภาพยนต์เรื่องนี้

ซึ่งหลักคมมีดของออกแคม สามารถอธิบายเพิ่มเติม ได้ดังนี้ ในกรณีของชุดข้อมูลสำหรับการฝึกฝนที่มีจำนวนน้อย เมื่อนำชุดข้อมูลนี้มาใช้ในการประมาณ เมทริกซ์ความแปรปรวนร่วมเกี่ยว (สมมติฐาน) ของ PCA ซึ่งมีขนาดมิติใหม่มาก (ความซับซ้อนสูงมาก) จะทำให้เกิด *การกระชับเกินสมควร* (overfitting) อันเป็นผลเนื่องมาจาก แบบจำลองที่มีความซับซ้อนมากเกินไปนั้น จะถูกรบกวนด้วย *สัญญาณรบกวนทางสถิติ* (statistical noise) ได้ง่าย เป็นที่น่าสังเกตว่า แบบจำลองที่ซับซ้อน นี้นั้นก็มีการอธิบายใน ทฤษฎี *การประมาณฟังก์ชัน* (function

approximation) ที่ซึ่งเราแทนรูปฟังก์ชันใดๆ ด้วย *ฟังก์ชันพหุนาม* (polynomial function) หลายอันดับ แต่ฟังก์ชันพหุนามอันดับสูงๆจะค่อนข้างอ่อนไหวต่อสัญญาณรบกวน ด้วยเหตุนี้ การประมาณ เมทริกซ์ความแปรปรวนร่วมเกี่ยว ของ 2DPCA ซึ่งมีขนาดมิติเล็กกว่ามาก (แบบ จำลองที่พื้นฐานกว่า) น่าจะสามารถจับเค้าโครงที่สำคัญได้ดีกว่า อันจะเป็นผลให้สมรรถนะในการทำนายปรากฏการณ์ที่ศึกษาดีกว่า

ปัญหา overfitting นั้นยังเป็นที่รู้จักกันดีทางทฤษฎีความน่าจะเป็นและการประมาณ และการรู้จำรูปแบบ โดยเฉพาะในหัวเรื่องที่เกี่ยวกับ *ทวิบทความเอนเอียงกับความแปรปรวน* (bias-variance dilemma) หรือ *การแลกเปลี่ยนระหว่างความเอนเอียงกับความแปรปรวน* (bias-variance trade-off) แต่ในบริบทที่แตกต่างออกไป ในการประมาณฟังก์ชันที่ใช้ฟังก์ชันพหุนามนั้น หากระบบที่เราสังเกตเป็นระบบเชิงเส้น แต่เราใช้ฟังก์ชันพหุนามที่มีอันดับสูงกว่าอันดับสอง มาใช้ในการประมาณระบบ และเรามีข้อมูลน้อยมากเกินกว่าจะสร้างสมการเพื่อแก้ไขหาค่าสัมประสิทธิ์ของชุดฟังก์ชันพหุนามทั้งหมดได้ ในกรณีนี้จะเกิดการ overfitting หรือ การประมาณที่มี ความแปรปรวน (variance) สูง เพราะฟังก์ชันพหุนามที่ใช้มีอันดับสูงกว่าระบบ แต่เมื่อใช้ข้อมูลในการฝึกฝนเยอะขึ้น ค่าสัมประสิทธิ์ของฟังก์ชันพหุนามอันดับสูงๆจะมีค่าเท่าศูนย์ คงเหลือเฉพาะค่าสัมประสิทธิ์ของฟังก์ชันพหุนามอันดับต้นที่เป็นเชิงเส้น ทำให้ overfitting หมดไป ในทางกลับกัน หากใช้ระบบที่ซับซ้อนมาก ซึ่งต้องใช้ฟังก์ชันพหุนามอันดับสูงๆมาใช้ในการประมาณจึงจะมีความเหมาะสม แต่เราใช้ชุดฟังก์ชันพหุนามที่มีอันดับต่ำเกินไป เราจะมีความเป็นไปได้ที่จะ เอนเอียง (bias) ไปยังกลุ่มข้อมูลบางกลุ่มมากเกินไป

ในการรู้จำใบหน้าซึ่งมีขนาดของมิติใหญ่มาก หากระบบจำแนกมีความซับซ้อนมาก (เหมือนกับ การใช้ชุดฟังก์ชันพหุนามอันดับสูงมาก) มาใช้ในการตัดสินคลาสของบุคคล ซึ่งเราจำเป็นจะต้องใช้ตัวอย่างในการฝึกฝนเป็นจำนวนมาก ซึ่งเป็นไปไม่ได้ในทางปฏิบัติในการรู้จำใบหน้า เพราะโอกาสที่จะเก็บภาพใบหน้าเป็นจำนวนมากไม่สามารถเป็นไปได้ เช่น เราอาจมีโอกาสเก็บภาพใบหน้าของผู้ก่อการร้ายได้เพียงไม่กี่ภาพใบหน้า เท่านั้น เป็นต้น ระบบจำแนกที่ซับซ้อนจะทำให้เกิด การจำข้อมูล หรือ มีค่าความแปรปรวนในการตัดสินใจสูง ในทางกลับกัน หากใช้ระบบจำแนกที่ซับซ้อนน้อย ก็มีความเป็นไปได้ที่เราจะเอนเอียงไปยังกลุ่มภาพใบหน้าบางกลุ่มมากเกินไป เราจึงควรสร้างระบบจำแนก และ ตัวสกัดบ่งต่าง ที่มีความซับซ้อนไม่มาก หรือน้อยจนเกินไป

นอกจากนี้ ปัญหา SSS หรือ overfitting ยังเป็นที่รู้จักดีในอีกชื่อว่า *คำสาบของมิติ* (curse of dimensionality) Richard Bellman (Bellman, 1961) ได้อธิบายถึงปัญหาที่เกิดขึ้นเมื่อปริมาตรเพิ่มขึ้นอย่างรวดเร็ว เมื่อจำนวนของแกนพิกัดเพิ่มขึ้นแก่ปริภูมิที่สนใจ สิ่งนี้เป็นอุปสรรคที่สำคัญในการวิเคราะห์ข้อมูลที่ขนาดมิติใหญ่มาก ซึ่งเป็นผลจากความเป็นจริงที่ว่า จากตำแหน่งซึ่งเดิมอยู่รวมกันเป็นท้องถิ่น จะไม่อยู่รวมกันต่อไป แต่จะกระจายเป็นวงกว้างเมื่อจำนวนมิติขยายใหญ่มากขึ้น หือกล่าวอีกนัยหนึ่ง คือ *ความเป็นมากศูนย์* (sparsity) จะเพิ่มในสัดส่วนแบบเอก

โปเนนเชียล ต่อ จำนวนตำแหน่งข้อมูลที่คงที่ โดยทั่วไป คำสาบของมิติ เป็นอุปสรรคที่สำคัญต่อ *การเรียนรู้ของเครื่อง* (machine learning) ในการเรียนรู้สภาวะของข้อมูลที่มีจำนวนตัวอย่างที่ใช้สำหรับฝึกฝนจำกัดบนปริภูมิตัวบ่งต่างที่มีขนาดมิติใหญ่ม

Bellman ต้องการชี้ให้เห็นว่า ปัญหาที่มีมีมิติใหญ่มากๆ จะต้องใช้สมมติฐานที่มีความซับซ้อนในการอธิบายปรากฏการณ์ และต้องใช้ข้อมูลปริมาณมากเพื่อใช้ในการสร้างสมมติฐานนั้น ซึ่งสามารถแปลความหมายสำหรับการรู้จำใบหน้าและเป้าหมายอัตโนมัติ ได้ดังนี้ เนื่องจากภาพใบหน้าและเป้าหมายประกอบด้วยพิกเซลจำนวนมาก (มิติมีขนาดใหญ่มาก) ระบบรู้จำ (แบบจำลองหรือสมมติฐาน) จึงถูกสาบให้ต้องใช้จำนวนภาพใบหน้าหรือเป้าหมายในการฝึกฝนเป็นจำนวนมาก (จำนวนภาพใบหน้าที่ใช้ฝึกฝนควรจะมีเป็นจำนวนเท่าของจำนวนพิกเซลของภาพ) เพื่อที่จะให้ระบบมีสมรรถนะที่สูงในการรู้จำ (ระบบมีความสามารถในการอธิบาย หรือจำแนกภาพใบหน้าหรือเป้าหมายได้อย่างถูกต้อง) ดังนั้น การที่เราใช้ ตัวบ่งลักษณะ เช่น 2DPCA ซึ่งคำนวณได้ง่ายและมีมิติที่น้อยกว่า ไม่เพียงแต่เราจะสามารถสร้างระบบรู้จำที่พื้นฐานกว่า ซึ่งทำให้สมมติฐานของระบบรู้จำที่สร้างขึ้นมีความซับซ้อนน้อยกว่า อันจะทำให้จำนวนภาพที่ต้องใช้ในการฝึกฝนมีจำนวนที่น้อยกว่า และ ระบบมีโอกาสที่จะทำนายได้ถูกต้องมากกว่าแล้ว เวลาที่ใช้ในการจำแนกคลาสยังจะน้อยกว่าอีกด้วย

นอกจากนี้ปัญหาสุดท้าย และ เป็นพื้นฐานที่สำคัญที่สุด คือการคำนวณ การแยกย่อยค่าเอกฐาน หรือ SVD ที่คอมพิวเตอร์จะไม่สามารถคำนวณได้ หาก เมทริกซ์ความแปรปรวนร่วมเกี่ยวมีขนาดใหญ่มาก

2DPCA แทนรูปภาพใบหน้าด้วยผลเฉลี่ยที่มีการถ่วงน้ำหนัก ใบหน้าลักษณ์ เช่นเดียวกับ PCA แตกต่างกันเฉพาะที่เมทริกซ์ความแปรปรวนร่วมเกี่ยว **G** ที่แตกต่างจาก **R** ของ PCA เราจะทำการแทนรูปชุดภาพใบหน้า เมทริกซ์ **A** ประกอบด้วยภาพใบหน้า x_i (อยู่ในรูปเมทริกซ์ภาพไม่มีการแปลงให้เป็นเวกเตอร์) วางต่อกันตามแนวแถวตั้งต่อเนื่องกันไป หรือ $A=[x_1,\cdots,x_M]$ หากกำหนดให้ $n\times m$ เป็นขนาดความกว้างและสูงของภาพใบหน้า และ M เป็นจำนวนตัวอย่างที่ใช้ในการฝึกฝน ดังนั้น เมทริกซ์ **A** จึงมีขนาด $n\times mM$ ซึ่งแตกต่างจาก PCA

ใน 2DPCA เมทริกซ์ความแปรปรวนร่วมเกี่ยวภาพ **G** ถูกนิยามให้เป็น

$$G = E\left[\left(A- E[A]\right)^T\left(A -E[A]\right)\right] \tag{9}$$

โดยที่ เมทริกซ์ **A** แทนเมทริกซ์รูปภาพใบหน้าในรูปแบบใหม่ และ $E[A]$ แทนค่าควาคาดหวังของเมทริกซ์ **A** จะเห็นได้ว่า เมทริกซ์ความแปรปรวนร่วมเกี่ยว **G** จะมีขนาดมิติเท่ากับ $n\times n$ ซึ่งมีขนาดเล็กกว่า เมทริกซ์ความแปรปรวนร่วมเกี่ยว ของ PCA มาก ในทางปฏิบัติ เมทริกซ์ความแปรปรวนร่วมเกี่ยว **G** ที่ประมาณได้ (estimated covariance matrix) จะสามารถคำนวณได้อย่างแม่นยำในกรณีของการฝึกฝนที่มีจำนวนตัวอย่างน้อย

เมทริกซ์ความแปรปรวนร่วมเกี่ยว **G** ที่ประมาณได้ ที่คำนวณมาจาก เอนเซมเบล ของข้อมูลภาพใบหน้าสำหรับฝึกฝนจริง จะเขียนได้ดังนี้

$$G = \frac{1}{M}\sum_{k=1}^{M}(A_k - \bar{A})^T(A_k - \bar{A}) \tag{10}$$

โดยที่ คือ $\bar{A}$ ค่าเฉลี่ยเมทริกซ์แทนรูปภาพใบหน้า และ $\bar{A} = \frac{1}{M}\sum_{k=1}^{M} A_k$

กำหนดให้ $\mathbf{\Phi}' = \left[e_1', \cdots, e_K'\right]$ เป็นเวกเตอร์เฉพาะเรียงลำดับจากใหญ่ที่สุด K เวกเตอร์เช่นกัน สำหรับภาพใบหน้าแต่ละภาพ x_l เวกเตอร์น้ำหนัก $\mathbf{a}_l'$ สามารถคำนวณได้จาก *การฉาย* (projecting) ภาพใบหน้าลงบนใบหน้าลักษณ์ จำนวน *K* ใบหน้า $\mathbf{\Phi}' = \left[e_1', \cdots, e_K'\right]$ บน K *ปริภูมิย่อย* (ปริภูมิย่อย)

$$\mathbf{a}_l' = \mathbf{\Phi}'\left(\overline{x_l} - \bar{A}\right) \tag{11}$$

เอาท์พุตที่ได้จากการฉายข้างต้น จะได้เป็นเมทริกซ์ขนาด $n \times K$ ซึ่งต่างจาก PCA ที่เอาท์พุตได้เป็นเวกเตอร์ที่มีขนาด $K \times 1$

กระบวนการจำแนกภาพวัตถุ ด้วย 2DPCA สามารถใช้ระเบียบ *ตัวจำแนกค่าใกล้ที่สุด* (nearest neighbor classifier) ในการจำแนก ซึ่งจำเป็นต้องใช้ระยะห่างในการจำแนก ในที่นี้ระยะห่างระหว่าง *เมทริกซ์บ่งต่าง* (feature matrix) สองเมทริกซ์ $\mathbf{a}_i' = \left[a_{i1}', a_{i2}', \cdots, a_{iK}'\right]$ และ $\mathbf{a}_j' = \left[a_{j1}', a_{j2}', \cdots, a_{jK}'\right]$ ซึ่งได้มาจากการฉาย ภาพวัตถุสองภาพ ลงบนเวกเตอร์เฉพาะของภาพวัตถุ ซึ่งในที่นี้ เวกเตอร์เฉพาะของภาพวัตถุ หมายถึงใบหน้าเฉพาะในกรณีของการรู้จำใบหน้า และ เป้าหมายเฉพาะ ในกรณีของการรู้จำเป้าหมายอัตโนมัติ ตามลำดับ โดยที่สมาชิกของแต่ละเมทริกซ์ เป็นแถวนอนที่ได้จากการฉายภาพวัตถุลงไปบนแต่ละเวกเตอร์เฉพาะของภาพวัตถุ

ระยะห่างระหว่าง เมทริกซ์บ่งต่างสองเมทริกซ์ ข้างต้น อาจกำหนดให้เป็น *มาตรระยะห่างชนิดบวก* (additive distance measure) ซึ่งสามารถเขียนขึ้นเป็นสมการ ได้ดังนี้

$$d(\mathbf{a}_i', \mathbf{a}_j') = \sum_{p=1}^{K}\left\|a_{ip}' - a_{jp}'\right\| \tag{12}$$

โดยที่ $\left\|a_{ip}' - a_{jp}'\right\|$ แทน ระยะห่างยุคลิด (Euclidean distance) ระหว่าง แต่ละ *เวกเตอร์บ่งต่างมุขสำคัญ* (principal feature vector) $\mathbf{a}_i'$ และ $\mathbf{a}_j'$ สมมติว่า เราใช้ตัวจำแนกแบบ 1-ตัวจำแนกค่าใกล้ที่สุด ที่ฝึกฝนด้วยชุดข้อมูลสำหรับฝึกฝน *M* ตัวอย่าง และ เมทริกซ์บ่งต่าง $\mathbf{a}_q'$ โดยที่ $q = 1, 2, 3, \cdots, M$ สามารถคำนวณได้จากการฉายแล้ว

เราสามารถทำนายคลาสของภาพวัตถุที่นำมาทดสอบ จากเมทริกซ์บ่งต่างที่คำนวณได้ $\mathbf{a}'_{test}$ ดังนี้ ถ้า

$$d(\mathbf{a}'_{test}, \mathbf{a}'_{w}) = \min_{q} d(\mathbf{a}'_{test}, \mathbf{a}'_{q}) \tag{13}$$

และ $\mathbf{a}'_{w} \in \omega_c$ แล้ว ภาพทดสอบ x_{test} จะถูกจำแนกให้อยู่ในคลาส c

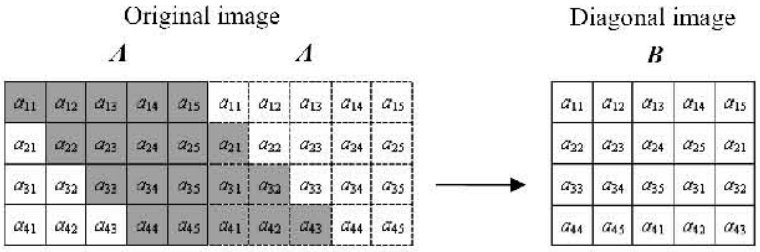
2.1.3 สองมิติรุ่นถัดมา

อย่างไรก็ดี หนึ่งในข้อด้อยของ 2DPCA ขนาดมิติของสัมประสิทธิ์ที่ใช้ในการแทนรูปภาพใบหน้านั้น จะมีขนาดใหญ่กว่า PCA มีงานจำนวนมากที่นำเสนอเพื่อปัญหาดังกล่าวนี้ เช่น การใช้ PCA ในขั้นที่สองต่อจากการทำ 2DPCA หรือ การนำเสนอ 2DPCA แบบมี *การฉายสองทาง* (bilateral-projection) (Zhang, n.d. & Hong, 2005) ที่ซึ่งมีการแยกย่อยเมทริกซ์ของเวกเตอร์เฉพาะออกเป็น เมทริกซ์ซ้าย-ขวา เมทริกซ์ของเวกเตอร์เฉพาะซ้าย-ขวาจถูกคำนวณอย่างเป็นอิสระต่อกัน โดยการคำนวณแบบวนซ้ำ ในการคำนวณ เมทริกซ์ของการฉายด้านขวาจะถูกคำนวณจากภาพที่ถูกสร้างคืนด้านซ้าย และ เมทริกซ์ของการฉายด้านซ้ายจะถูกคำนวณจากภาพที่ถูกสร้างคืนด้านขวา และทำการคำนวณแบบวนซ้ำต่อเนื่องไป คำตอบที่ได้จะเป็นคำตอบ *เล็งเลิศท้องถิ่น* (local optimum)

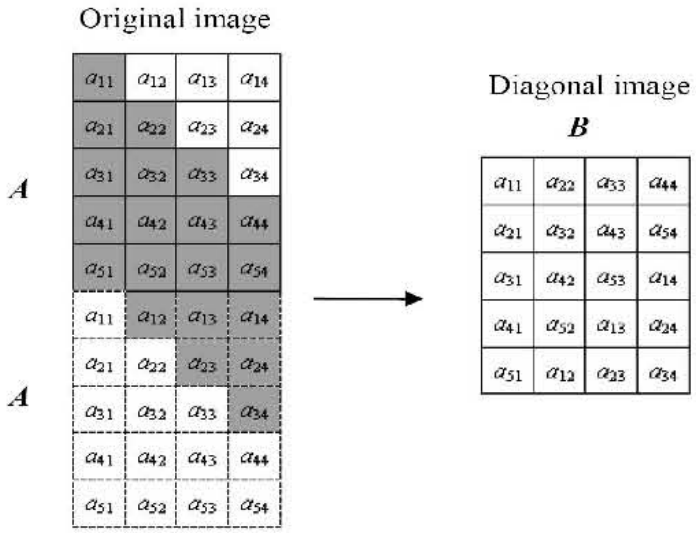
เป็นที่น่าสังเกตว่าระเบียบวิธี 2DPCA ที่นำเสนอในตอนแรกนั้น เราสามารถพิจารณาให้เห็นว่าเป็นการประมวลผลในลักษณะ *เชิงแถวนอน* (row-based) ซึ่งเราสามารถปรับให้การประมวลผลเป็นในลักษณะ เชิงแถวตั้ง (column-based) ได้ไม่ยาก (Sanguansat, 2006) ส่วน 2DPCA แบบการฉายสองทาง เป็นการประมวลผลที่คำนึงถึงข่าวสารของการแทนรูปทั้งในเชิงแถวนอน-ตั้ง หรืออีกนัยหนึ่ง 2DPCA ที่เน้นข่าวสารในแนวแถวตั้ง จะสกัดตัวบ่งต่างที่มีความสามารถในการคัดแยกคลาส ตามโครงสร้างของในแนวตั้งใบหน้า ทั้งนี้เป็นไปตามสมมติฐานที่ว่า ภาพควรจะมี *ความราบเรียบ* (smoothness) ในระดับหนึ่ง เราจึงสามารถละเลยข้อมูลในแนวระนาบราบได้ ดังนั้นเราจึงสามารถทำการลดรูปการวิเคราะห์ ด้วยการหา ความสัมพันธ์ร่วมเกี่ยวในระดับแถวนอนของชุดข้อมูลฝึกฝน สำหรับ 2DPCA แทนการคำนวณหา ความสัมพันธ์ร่วมเกี่ยวในระดับพิกเซลของชุดข้อมูลฝึกฝน สำหรับ PCA ในขณะที่ 2DPCA เน้นข่าวสารในแนวแถวตั้ง และ ตั้งใจที่จะสกัดตัวบ่งต่างที่มีความสามารถในการคัดแยกคลาสตาม โครงสร้างของภาพตามแนวตั้งของใบหน้า ข่าวสารที่สำคัญที่เกี่ยวกับโครงสร้างใบหน้าบางส่วนจึงถูกลดทอนความสำคัญลง ซึ่งเป็นผลเนื่องมาจากการสะสมความแปรปรวนในทิศทางนั้น ยกตัวอย่าง เช่น สำหรับ PCA โครงสร้างของตาจะถูกแทนรูปในระดับพิกเซล ในขณะที่ สำหรับ 2DPCA โครงสร้างของตาจะถูกแทนรูปอย่างสะสมตลอดแนวระดับของตา (หัวตาซ้าย ตาซ้าย หางตาซ้าย หัวตาขวา ตาขวา และ หางตาขวา)

อย่างไรก็ดี ความไม่ราบเรียบ ของโครงสร้างใบหน้า เป็นข่าวสารที่สำคัญในการรู้จำใบหน้า เช่น ความคมเข้มของคิ้ว และ ความสัมพันธ์ของคิ้วกับตา ของบุคคลบางคน เป็นต้น ด้วยเหตุนี้ จึง

มีการนำเสนอการคำนวณ 2DPCA โดยการลดปริมาณข่าวสารในบางทิศทาง แต่คงไว้ซึ่งข่าวสารในบางทิศทางไว้ ซึ่งสามารถเรียกว่า *2DPCA แนวขวาง* (diagonal 2DPCA) ซึ่งคำนวณจากการสร้างเมทริกซ์ความแปรปรวร่วมเกี่ยว ที่เกิดจากการแปลงภาพใบหน้าให้อยู่ในลักษณะ *แนวขวาง* (diagonal) ดังแสดงในรูปที่ 2.3 และ 2.4 เนื่องจาก แถวนอน (แถวตั้ง) ของภาพใบหน้าที่ถูกแปลงทางขวาง ได้ปรับเปลี่ยนปริมาณข่าวสารของภาพใบหน้าในแนวนอนและแนวดิ่งจากภาพใบหน้าดั้งเดิม ผลของปริมาณข่าวสารที่เปลี่ยนไปนี้ทำให้อาจมีข่าวสารที่สำคัญของโครงสร้างใบหน้าปรากฏอยู่ ไม่ถูกเฉลี่ย (สะสม) หายไป ภาพใบหน้าที่ถูกแปลงทางขวางได้แสดงไว้ ดังในรูปที่ 2.5



รูปที่ 2.3 ระเบียบวิธีในการสร้างภาพใบหน้าทางขวาง กรณีจำนวนแถวตั้ง (คอลัมน์) มากกว่าจำนวนแถวนอน



รูปที่ 2.4 ระเบียบวิธีในการสร้างภาพใบหน้าทางขวาง กรณีจำนวน แถวตั้ง (คอลัมน์) น้อยกว่าจำนวนแถวนอน

รูปที่ 2.5 ตัวอย่างภาพใบหน้าที่ถูกแปลงทางขวาง

เป็นที่น่าสังเกตว่า เราสามารถขยายแนวความคิดเกี่ยวกับ 2DPCA แนวขวาง ไปสู่ 2DPCA ที่สามารถระบุทิศทางได้ โดยการแปลงแนวขวางของภาพใบหน้าที่ถูกแปลงแนวขวางแล้วอีกครั้งหนึ่ง ก็จะได้ 2DPCA ใหม่ หากเราการแปลงแนวขวางซ้ำต่อเนื่องไปเรื่อยๆ เราจะสามารถสร้าง 2DPCA ขึ้นใหม่ได้ในทิศทางต่างๆ แต่ละทิศทาง 2DPCA ในลักษณะนี้สามารถเรียกอีกอย่างหนึ่งว่า *การวิเคราะห์ส่วนประกอบมุขสำคัญระบุทิศทาง* (directional two-dimensional principal component analysis: directional 2DPCA) หรือ 2DPCA ระบุทิศทาง

ทีนี้เราก็จะได้ 2DPCA ที่สามารถคำนวณสหสัมพันธ์ระหว่างแถวข้อมูลของตัวเอง (ดูรายละเอียดของคำอธิบายเพิ่มเติม เกี่ยวกับ ความสัมพันธ์ของความแปรปรวนร่วมเกี่ยวระหว่าง PCA กับ 2DPCA ในบทย่อยต่อไป) ซึ่งจะให้ตัวบ่งต่างที่บรรจุข่าวสารที่เป็นประโยชน์ต่อการรู้จำใบหน้าที่ต่างกัน หากเราใช้ชุดของตัวบ่งต่างเหล่านี้สร้างชุดของระบบรู้จำ แต่ละระบบรู้จำจะเน้นข่าวสารในการรู้จำที่ต่างกัน ซึ่งจะให้การตัดสินใจที่ต่างๆกัน หากนำผลการตัดสินใจเหล่านี้มาร่วมพิจารณาเพื่อหาการตัดสินใจในขั้นสุดท้าย โดยปกติจะให้ผลของการตัดสินใจสุดท้ายที่ดีกว่า เหมือนกับแนวความคิดทางประชาธิปไตย ที่ต้องการความเห็นของส่วนรวม หรือ ตามทฤษฎีความน่าจะเป็น ที่ความแปรปรวนในการตัดสินใจจะลดลง ทำให้ผลผิดพลาดของการประมาณการตัดสินใจลดลง ระบบรู้จำในลักษณะนี้จะเรียกว่า *ระบบจำแนกแบบหลายตัว*

2.1.4 สองมิติไขว้

สำหรับ PCA ความแปรปรวนร่วมเกี่ยว เป็นดัชนีในการวัดความเข้มของสหสัมพันธ์ที่มีต่อกันของคู่พิกเซล การรู้จำใบหน้า เป็นปัญหาการรู้จำที่มีมิติขนาดใหญ่มาก เหนือการนั้นเป็นปัญหาที่ยากในการหาตัวอย่างสำหรับฝึกฝน ด้วยเหตุนี้ โอกาสที่จะประมาณเมทริกซ์ความแปรปรวนร่วมเกี่ยวให้แม่นยำจึงเป็นไปได้ยาก 2DPCA ได้ถูกเสนอเพื่อนำมาใช้ในการแทนรูปภาพใบหน้า โดยการคำนวณฟังก์ชันฐานหลักชนิดปรับตัวได้จากชุดของข้อมูลภาพใบหน้าสำหรับการฝึกฝน 2DPCA สร้างเมทริกซ์ความแปรปรวนร่วมเกี่ยวที่ต่างจาก PCA เพื่อใช้สำหรับวัดความความเข้มของสหสัมพันธ์ที่มีต่อกันของแต่ละ แถวนอน (แถวตั้ง หรือ คอลัมน์) ของภาพใบหน้า มองให้ลึกลงไป จะเห็นว่า ความสัมพันธ์ของพิกเซลในแต่ละคู่ของแถวของภาพได้ถูกสะสมไว้ในเมทริกซ์ความแปรปรวนร่วมเกี่ยวที่เกิดขึ้นใหม่นี้ ด้วยเหตุนี้ ข่าวสาร (สหสัมพันธ์ของบางคู่

พิกเซล) บางส่วนได้ถูกทำลายไป เพื่อให้สหสัมพันธ์ที่สำคัญบางส่วนคงอยู่ คณะผู้วิจัย (Sanguangsat et al., 2006) ได้นำเสนอทางเลือกใหม่ ที่เรียกว่า *2DPCA ไขว้* (cross-2DPCA) ขึ้น จากการวิจัย คณะผู้วิจัยพบว่า เมทริกซ์ความแปร ปรวนร่วมเกี่ยวของ 2DPCA เกิดจากการคำนวณสหสัมพันธ์ระหว่างเมทริกซ์ชุดภาพใบหน้าฝึกฝน กับ ตัวเอง

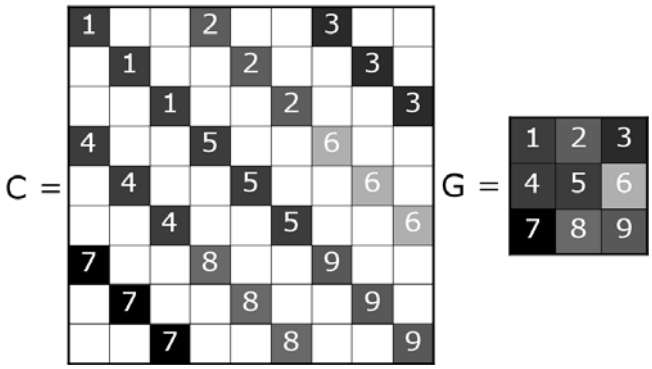
ในขณะที่ 2DPCA แนวขวาง ก็เกิดจากการคำนวณหาสหสัมพันธ์ระหว่างเมทริกซ์ชุดภาพใบหน้าฝึกฝนที่ถูกแปลงแนวขวาง กับตัวเอง ดังนั้น หากเราทำการคำนวณหาสหสัมพันธ์ที่เกิดจากการเข้าคู่ระหว่างทิศทางของแนวของภาพที่แตกต่างกัน เราจะสามารถเข้ารหัสข่าวสารที่เป็นประโยชน์ต่อการรู้จำไว้ในเมทริกซ์ความแปรปรวนร่วมเกี่ยวใหม่ ที่พัฒนาขึ้น

โดยการนิยามให้ *เมทริกซ์ความแปรปรวนร่วมเกี่ยวไขว่* (cross-covariance matrix), **G** เป็นไปดังนี้

$$G_{BA} = E[B^T A] \tag{14}$$

โดยที่ เมทริกซ์ ***A*** เป็นเมทริกซ์ของชุดภาพใบหน้า x_i (อยู่ในรูปเมทริกซ์ภาพไม่มีการแปลงให้เป็นเวกเตอร์) วางต่อกันตามแนวแถวตั้งต่อเนื่องกันไป หรือ $A=[x_1,\cdots,x_M]$ และ ***B*** เป็นเมทริกซ์ของชุดภาพที่สร้างขึ้นใหม่ โดยการแปลงโดยการเลื่อนตำแหน่งพิกเซล รูปที่ 2.6 แสดง 100 ตัวอย่างของ เมทริกซ์ ***B*** ที่เกิดจากการเลื่อนพิกเซลทั้งในแนวนอนและแนวตั้งของภาพใบหน้าของ เมทริกซ์ ***A***

รูปที่ 2.6 ตัวอย่างของภาพใบหน้าที่เกิดการเลื่อนทั้งในแนวนอนและแนวตั้ง



รูปที่ 2.7 สมาชิกของเมทริกซ์ความแปรปรวนร่วมเกี่ยวของ PCA และ 2DPCA

ความสัมพันธ์ระหว่าง ความแปรปรวนร่วมเกี่ยว PCA 2DPCA และ 2DPCA ไขว้

เบื้องต้น ความแปรปรวนร่วมเกี่ยวของ 2DPCA สามารถคิดค้นมาจาก ความแปรปรวนร่วมเกี่ยวของ PCA ได้ดังสมการต่อไปนี้

$$G(i,\ j)\ =\sum_{k=1}^{m}C\left(m(i-1)+k,m(j-1)+k\right) \qquad (15)$$

โดยที่ $G(i,\ j)$ เป็นสมาชิกแถวนอนที่ i และ แถวตั้งที่ j ของเมทริกซ์ความแปรปรวนร่วมเกี่ยวของ 2DPCA และ m คือขนาดความสูงของภาพใบหน้า เราได้แสดงการสาธิตความ สัมพันธ์ดังกล่าว ดังตัวอย่างต่อไปนี้ สมมติให้เมทริกซ์ภาพใบหน้ามีขนาด 3×3 เมื่อทำให้ภาพใบหน้าให้เป็นเวกเตอร์ ที่มีขนาดมิติเท่ากับ 9 เราจะได้เมทริกซ์ความแปรปรวนร่วมเกี่ยวของ PCA ที่มีขนาดมิติเท่ากับ 9×9 ส่วนเมทริกซ์ความแปรปรวนร่วมเกี่ยวของ 2DPCA จะมีขนาดมิติเท่ากับ 3×3 จากสมการที่ (15) และ รูปที่ 2.7 สมาชิกแต่ละสมาชิกของเมทริกซ์ G ที่มีหมายเลขสมาชิกกำกับ เกิดจากผลรวมของสมาชิกที่มีหมายเลขสมาชิกตรงกันที่อยู่ในเมทริกซ์ C จากเหตุผลข้างต้น จะเห็นได้ว่า สมาชิกของเมทริกซ์ความแปรปรวนร่วมเกี่ยวของ 2DPCA มีข่าวสารแค่ 1 ใน m ของเมทริกซ์ของ PCA หรือ อีกนัยหนึ่ง สมาชิกของเมทริกซ์ความแปรปรวนร่วมเกี่ยวของ 2DPCA เกิดจากการสะสมค่าของสมาชิกจำนวน m ตัวของเมทริกซ์ของ PCA แม้ทว่าข่าวสารที่สำคัญของภาพใบหน้าจะสูญเสียไปบ้างเมื่อทำการสกัดตัวบ่งต่างด้วย 2DPCA หรือ อีกนัยหนึ่ง คือในการประมาณเมทริกซ์ความแปรปรวนร่วมเกี่ยวของ 2DPCA จะเกิดความเอนเอียงเกิดขึ้น แต่ข้อได้เปรียบของ 2DPCA อยู่ตรงด้านความแม่นยำในการประมาณเมทริกซ์ความแปรปรวนร่วมเกี่ยวในกรณีที่มีจำนวนตัวอย่างในการฝึกฝนน้อย เพื่อหลีกเลี่ยงปัญหา SSS หรือ คมมีดของออกแคม หรือ คำสาปของมิติ ก็ยังทำให้ 2DPCA เป็นระเบียบวิธีที่น่าสนใจกว่า PCA

จากสมการที่ (15) เราสามารถขยายความ ความสัมพันธ์ระหว่าง เมทริกซ์ความแปรปรวนร่วมเกี่ยวของ PCA กับ เมทริกซ์ความแปรปรวนร่วมเกี่ยวไขว้ ได้ดังสมการต่อไปนี้

$$G(i,\ j)\ =\sum_{k=1}^{m}C\left(\ f\left(m(i-1)+k\right),m(j-1)+k\right) \qquad (16)$$

โดยที่

$$f(x)=\begin{cases} x+L-1, & 1\le x\le mn+L+1 \\ x-mn+L-1, & mn-L+2\le x\le mn \end{cases} \quad (16)$$

และ $1\le L\le mn$

ตามทฤษฎีความน่าจะเป็น ค่าผิดพลาดเฉลี่ยกำลังสองในการประมาณ ประกอบด้วยผลบวกของกำลังสองของความเอนเอียง กับ ความแปรปรวน เนื่องจาก 2DPCA ไขว้ สามารถคำนวณได้ หนึ่งตัวบ่งต่างต่อทิศ โดยแต่ละตัวบ่งต่างอาจมีความเอนเอียงในการประมาณอยู่บ้างเมื่อเทียบกับ PCA หากว่า ความเอนเอียงของแต่ละตัวบ่งต่างที่สกัดจาก 2DPCA ไขว้ มีค่าน้อย ตามทฤษฎีความน่าจะเป็น เมื่อเรานำตัวบ่งต่างเหล่านี้ใช้ร่วมกัน ความแปรปรวนลงจะลดลงโดยประมาณ *p* เท่าของความแปรปรวนของตัวบ่งต่างแต่ละตัว โดยที่ *p* คือจำนวนตัวบ่งต่างที่สกัดด้วย 2DPCA ไขว้ ด้วยเหตุนี้ การรู้จำด้วย คณะของ 2DPCA ระบุทิศทาง หรือ 2DPCA ไขว้ ควรจะได้ผลดีในการตัดสินใจขั้นสุดท้ายดีขึ้น

2.1.5 ความเป็นไปได้ทางชีวภาพของมองเห็น

ในบทย่อยนี้ เราจะกล่าวถึง PCA และ 2DPCA และ 2DPCA รุ่นถัดมา กับความเป็นไปได้ทางชีวภาพของการมองเห็น โดยเราจะชี้ให้เห็นโดยลำดับ ดังนี้ อันดับแรกเกี่ยวข้องกับการสกัดความหมายจากการมองเห็น ด้วยการประมวลผลของกลุ่มเซล *ปมประสาท* (ganglion) ที่ได้รับจากจอประสาทตา ซึ่งสามารถชี้ให้เห็นว่า PCA มีการทำงานที่สามารถเทียบเคียงได้ อันดับที่สอง จากการศึกษางานทางด้าน *ประสาทวิทยา* (neurological) พบว่า (receptive field) มีผลตอบสนองต่อทิศทาง ซึ่งเป็นแนวคิดที่สนับสนุนการพัฒนาต่อยอด 2DPCA เป็นชนิดระบุทิศทาง เพื่อให้สามารถเข้ารหัสข่าวสารในแต่ละทิศทางได้ รวมถึงขยายไปถึงการพัฒนา 2DPCA ไขว้ เมื่อคำนึงถึงลักษณะการทำงานของปมประสาทที่ทำงานร่วมกันเป็นกลุ่มอย่าง *อสัณฐาน* (irregular) ชี้นำให้เกิดแนวคิดในการ ชุดของ 2DPCA ชนิดระบุทิศทาง และ แบบไขว้ ไว้กับระบบจำแนกแบบหลายตัว

โดยสรุป หลักการของการนำกลุ่มของ 2DPCA ไขว้ มาใช้ในระบบจำแนกแบบหลายตัว โดยการบูรณาการ ทวิบทของความเอนเอียงกับความแปรปรวน มาใช้กับการรู้จำ ซึ่งราย ละเอียดจะได้กล่าวถึงในบทต่อไป

หลักฐานทางการวิจัยอันดับแรกที่เกี่ยวข้อง กับ ความเป็นไปได้ทางชีวภาพของการมองเห็น ซึ่งใช้ PCA เป็นตัวสกัดตัวบ่งต่าง หนึ่งในงานวิจัยที่น่าสนใจได้แก่งานของ Weblin (Werblin, 2007) ที่อธิบาย การสกัดความหมายจากการมองเห็นด้วยการประมวลผลของกลุ่มเซล *ปมประสาท* (ganglion) ต่อสัญญาณภาพที่ได้รับจากจอประสาทตา Werblin เสนอว่า การประมวลผลประกอบด้วยการแทนรูปทัศนียภาพธรรมชาติ (natural scene) โดยการใช้ ภาพ

รูปที่ 2.8 การประมวลผลภาพด้วยชุดเซลของปมประสาท โดยการแทนรูป 12 ภาพ (ในภาพแสดงภาพแทนรูปบางส่วน) สำหรับภาพสี่ภาพในแนวนอนแสดงผลตอบสนองต่อเวลา ภาพที่ได้ดัดแปลงมาจากวารสาร Scientific American (Werblin, 2007)

สำหรับการแทนรูป จำนวน 12 ภาพ การทดลองของเขาประกอบด้วย การจำลองด้วยโปรแกรมที่ทำงานร่วมกับ ชิปจอประสาทตาคอมพิวเตอร์ และ ทำการยืนยันผลการจำลองของโปรแกรมด้วย การวัดปฏิกิริยาตอบสนองจากจอประสาทตาของกระต่าย ในขณะที่ดูที่ภาพใบ หน้าของผู้พูด ผลจากการทดลองที่ได้แสดงดังในรูปที่ 2.8 จากรูปจะเห็นว่า การแทนรูปเสมือนผ่านตัวกรองตัวที่ 1 (สีส้ม ภาพในวงกลมที่อยู่บนสุด) ซึ่งดูเหมือนจะทำหน้าที่ตรวจจับลักษณะบ่งต่างเฉพาะส่วนที่เป็น *ขอบ* (edge) ส่วนการแทนรูปอีกส่วนหนึ่ง (สีม่วง ภาพที่สามจากด้านบน) จะทำหน้าที่สกัดเงาที่อยู่ใต้ตาและจมูกส่วนการแทนรูปอีกรู)หนึ่ง (สีน้ำตาลอ่อน ภาพที่สี่จากด้านบน) จะเน้นลักษณะบ่งต่างในส่วนที่จ้าที่สุดของภาพใบหน้า การแทนรูปบางส่วนอาจให้ข่าวสารเกี่ยวกับ *ลวดลายบนพื้นผิว* (texture) นอกจากนี้ ในรูปที่ 2.8 ยังได้แสดงถึง *ผลรวมของการแทนรูป* (combined representation) ด้วยภาพแทนรูปทั้ง 12 ภาพ ซึ่งเปลี่ยนตามเวลา ที่ 1 2 3 และ 4 ตามลำดับ

เป็นที่น่าสังเกตว่า ผลตอบสนองของการมองเห็นด้วยจอประสาทตาที่ประมวลผลด้วยชุดของปมประสาท จะเหมือนการดูภาพยนตร์ (มีข่าวสารตามเวลาอยู่ด้วย) ซึ่งต่างจากรู้จำที่งานวิจัยที่กำลังดำเนินการอยู่นี้ หากแต่ว่าในอนาคต เราสามารถขยายขอบเขตงานวิจัยให้ควบรวมข่าวสารทางเวลาได้ซึ่งจะทำให้ความแม่นยำในการรู้จำของระบบของเราดีขึ้น

วัตถุประสงค์ที่เราได้กล่าวถึงความเป็นไปได้ทางชีวภาพของการมองเห็นข้างต้นนี้ เพื่อชี้ให้เห็นถึงความคล้องจ้องกับงานวิจัยที่ดำเนินการอยู่ หากเราพิจารณารูปที่ 2.2 อีกครั้งหนึ่ง จะเห็นว่าตัวอย่างใบหน้าลักษณ์ (eigenface) ที่เรียงตามขนาดของ ค่าเฉพาะ (eigenvalue) นั้น ในใบหน้าลักษณ์บางใบหน้าจะแสดงลักษณะ (แทนรูป) ความราบเรียบของผิวหน้า ใบหน้าลักษณ์บางใบหน้าแสดงเงาใต้จมูก เป็นต้น ซึ่งคล้ายคลึงกับผลตัวกรองที่ปรากฏในงานทดลองของ Werblin และ เช่นเดียวกับ PCA ที่ให้ผลรวมแบบถ่วงน้ำหนักเฉพาะจะสามารถแทนรูปภาพใบหน้าจริงเหมือนผลการทดลองข้างต้นได้ เหมือนดังข้อสรุปของ Werblin ที่ว่า **การประมวลผลของชุดของปมประสาทจากสัญญาณที่ได้รับจากจอประสาทตาเป็นผลจากผลรวมของการแทนรูป**

อย่างไรก็ดี ความเป็นไปได้ทางชีวภาพในการมองเห็นของ PCA ที่ได้กล่าวมาข้างต้น ยังไม่มีการใช้ควบรวมข่าวสารเฉพาะของวัตถุที่คงอยู่ในแต่ละทิศทาง ต่อไปนี้ เราจะกล่าวถึงแรงบันดาลใจที่จะชี้ให้เห็นที่ความสำคัญของการสกัดข่าวสารที่ขึ้นกับทิศทาง เพื่อพัฒนาให้การประยุกต์ให้ PCA มีความน่าจะเป็นไปได้ทางชีวภาพมากขึ้น

เพื่อโน้มน้าวให้เห็นถึงความเหมาะสมในการใช้ 2DPCA แบบระบุทิศทาง หรือ 2DPCA แบบไขว้ ใน *การรู้จำวัตถุ* (object recognition) เราจะทำความเข้าใจโดยผ่านข้อคิดที่น่าสนใจเกี่ยวกับความสามารถในการรู้จำเมื่อวัตถุมีการหมุนในแนวระนาบราบและดิ่ง ดังต่อไปนี้

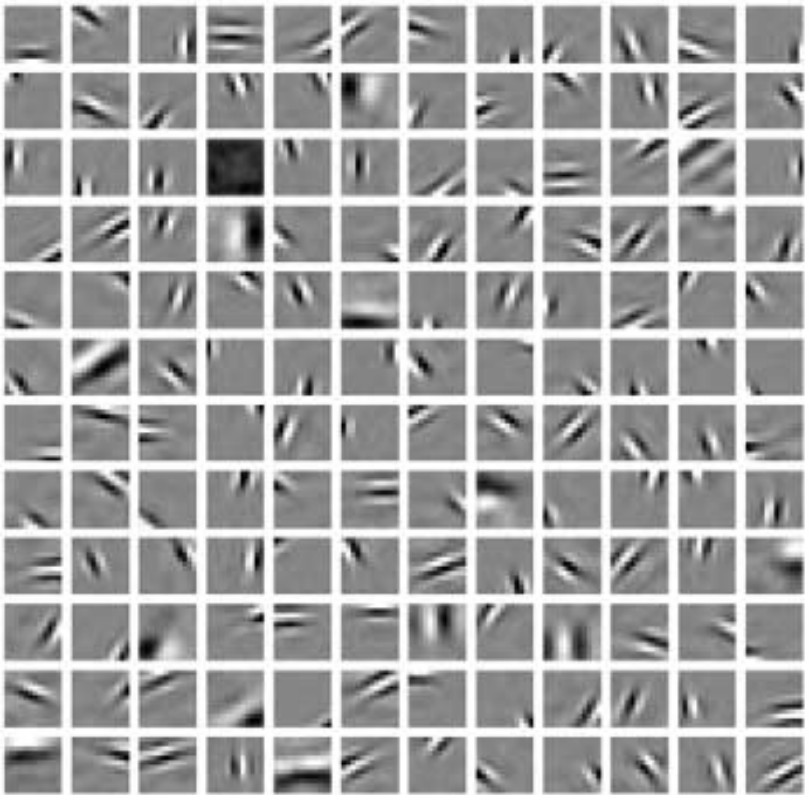
จากการศึกษางานวิจัยทาง *ประสาทวิทยา* (neurological) เกี่ยวกับความเป็นไปได้ทางชีวภาพในการมองเห็น (Roger, 1985) พบว่า

> *"ทิศทางที่เราส่วนใหญ่สามารถรับรู้ความรู้เชิงวัตถุวิสัยได้อย่างทันใด คือ ทิศทางที่นิยามให้อยู่ในแนวดิ่งตามแรงโน้มถ่วงของโลก ทิศทางที่สำคัญที่สุดที่ทำให้เราตั้งตรงและคงอยู่ คือทิศทางที่โดยปกติตั้งฉากกับแนวราบ (horizontal) ของพื้นดิน"*

หรือที่ งานวิจัย อีกชิ้นหนึ่ง (Edelman, 1992) ที่กล่าวว่า

> *"วัตถุสามารถคงทนต่อทิศ* (orientation) *ทางแนวราบ ต่อ ระนาบแนวดิ่ง ได้มากกว่าสามเท่า ที่ซึ่งการปรับทิศในระนาบแนวดิ่งเพียงเล็กน้อยจะมีผลทำให้การรู้จำลดลงในอัตราส่วนที่รวดเร็ว"*

จากข้อความข้างต้น เราสามารถสรุปได้ว่า 2DPCA แบบแถวตั้ง (column-based) จะมีความสามารถในการรู้จำมากกว่า 2DPCA แบบแถวนอน (row-based) ศึกษารายละเอียดเพิ่มเติมในบทต่อไป



รูปที่ 2.9 ภาพเล็กแต่ละภาพ แสดงกลุ่มประสาทสัมผัสสำหรับแบบ จำลองของเซลประสาทในการมองเห็น (Olshuasen, 2004)

จากข้อคิดข้างต้น เราสามารถขยายจากแนวระนาบราบ และ แนวตั้งฉาก ให้มีทิศทางมากขึ้น โดยอ้างถึงงานวิจัยชั้นนำของ Olshausen (Olshausen, 2004) ที่เสนอว่า *กลุ่มประสาทสัมผัส* (receptive field) ที่อยู่ที่สมองส่วนหน้า จะทำการประมวลผลสัญญาณจากจากภาพที่ตกลงที่จอประสาทตา ด้วย *การแทนรูปแบบมากเลขศูนย์* (sparse representation) ระบบประสาทใช้เซลประสาทจำนวนเพียงไม่มากนักในการแทนรูปในลักษณะนี้ ทำให้ขนาดของหน่วยความจำมีขนาดเล็กและมีประสิทธิภาพในการเข้ารหัสการมองเห็น นอกจากนี้ ยังมีงานวิจัยอีกเป็นจำนวนมากที่พบว่า มีเซลปมประสาทที่ไวต่อทิศทางในจอประสาทตา และสมองรับรู้ส่วนหน้าของสัตว์เลี้ยงลูกด้วยนมส่วนใหญ่ รูปที่ 2.9 แสดงกลุ่มประสาทสัมผัสที่สังเคราะห์ได้จากการฝึกฝนด้วยภาพจริงจำนวนมาก กลุ่มประสาทสัมผัส ที่ได้นี้ จะมีทิศ (orientation) มีลักษณะใกล้ถิ่นทางตำแหน่ง (spatially localized) และ เป็นตัวกรองผ่านกลาง (แบนด์พลาส)

จากเหตุผลข้างต้นนี้ เราจึงเชื่อได้ว่า 2DPCA ที่สามารถระบุทิศทางได้ น่าจะมีความสามารถในการรู้จำที่แม่นยำมากขึ้น และเนื่องจากปมประสาทจะร่วมกันทำงานในการรู้จำโดยแต่ละเซลประสาทจะเข้ารหัสข่าวสารที่แตกต่างกันตามทิศของตนเอง และนี้คือเหตุผลสุดท้ายที่เรานำมาใช้สนับสนุน ในการใช้ 2DPCA แบบระบุทิศทาง หรือ 2DPCA แบบไขว้ จำนวนหลายตัวมาเพื่อสร้างระบบจำแนกแบบหลายตัว ในการตัดสินวัตถุร่วมกัน

2.2 การสร้างคืนภาพความละเอียดสูงยวดยิ่ง

เนื่องจากข้อจำกัดทางกายภาพ ทำให้คุณภาพความละเอียดของภาพที่ได้ในการใช้งานต่างๆเกิดข้อจำกัด ระบบภาพเหล่านี้จะเกิด *สัญญาณแฝง* (aliased) หรือ ถูกอัตราสุ่มสัญญาณต่ำกว่าที่ควรจะเป็น หากจำนวนแถวของตัวรับไม่หนาแน่นพอ ซึ่งค่อนข้างพบได้บ่อยในอุปกรณ์ภาพอินฟาเรด และ กล้องที่ใช้ charge coupled device (CCD) มีงานวิจัยหลายชิ้นที่พัฒนาเทคนิคทางตรง (direct) และแบบวนซ้ำ (iterative) สำหรับความละเอียดสูงยวดยิ่ง หรือ การสร้างคืนภาพความละเอียดสูง (high-resolution: LR) ที่ปราศจากสัญญาณแฝง จากภาพความละเอียดต่ำ (low-resolution: LR) ที่มีสัญญาณแฝง สำหรับระเบียบวิธีทางตรงจะทำการสกัดองค์ประกอบความถี่สูงจากข้อมูลความถี่ต่ำที่มีอยู่ในเฟรม LR ด้วยปริภูมิฟูเรียร์ หรือ เราอาจใช้แนวทางการรวมกัน ของ การประมาณในช่วง-การบูรณะ (interpolation-restoration) หรือ การพิจารณาปัญหาเป็นปัญหาของการสุ่มสัญญาณแบบมีนัยทั่วไปของข้อมูลแบบรายคาบ นอกจากนี้ยังพิจารณาการสร้างคืน แบบ *ปัญหาผกผัน* (inverse problem) ที่อาจจำเป็นต้องใช้การแก้ปัญหาแบบวนซ้ำ ที่นิยมใช้กับปัญหาผกผันที่มีขนาดใหญ่

ในการสร้างคืนด้วยการคำนวณแบบวนซ้ำนั้น เราจำเป็นต้องทราบพารามิเตอร์ต่างๆ เหล่านี้ คือ *ตัวดำเนินการในการทำพร่า* (blur operator) การแปลงระนาบ *สัมพรรค* (affine transform) และ ตัวดำเนินการในการลดอัตราการชักตัวอย่าง โดยผ่านการลงทะเบียน (registration) ทางภาพ เพื่อสร้างเมทริกซ์ที่กระทำต่อภาพความละเอียดสูงยวดยิ่งแล้วจะได้ชุดของภาพความละเอียดต่ำ โดยปกติเราจะทำการแก้ปัญหาผกผัน ด้วยการย้ายข้างเมทริกซ์ที่กระทำต่อ HR ไปด้านของชุดภาพความละเอียดต่ำ เราก็จะได้ภาพความละเอียดสูงยวดยิ่ง แต่เนื่องจากสมการที่ตั้งขึ้นนี้มีขนาดใหญ่มาก ในการแก้ปัญหาจึงนิยมใช้ระเบียบวิธีแบบวนซ้ำ

โดยสรุป ระเบียบวิธีที่ใช้ในการสร้างคืนภาพความละเอียดสูงยวดยิ่ง แบ่งออกได้เป็น 2 แนวทาง ได้แก่ ทางความถี่ (frequency) และ ทางปริภูมิตำแหน่ง (space) ในงานวิจัยนี้จะเน้นการสร้างคืนทางปริภูมิตำแหน่งด้วยระเบียบวิธีแบบวนซ้ำ

2.2.1 แบบจำลองภาพความละเอียดสูงยวดยิ่ง

คณิตศาสตร์ในการคำนวณเชิงเลขตามกรอบของการสร้างคืนภาพความละเอียดสูงยวดยิ่ง (Nguyen et al., 2001) กำหนดให้ความสัมพันธ์ระหว่าง ภาพความละเอียดสูง (high resolution: HR) กับ ภาพความละเอียดต่ำ (low resolution: LR) ที่สามารถเขียนอยู่ในรูปเมทริกซ์ เป็นไปดังต่อไปนี้

$$\mathbf{f}_k = D_k B_k E_k \mathbf{x} + \mathbf{n}_k, \quad 1 \le k \le p \tag{1}$$

โดยที่ p เป็นจำนวนเฟรม $\mathbf{f}_k$ และ $\mathbf{x}$ เป็นเวกเตอร์ที่ได้จาก เฟรมของภาพ LR ที่ k^{th} และ ภาพ HR ที่ถูกจัดเรียงตามลำดับแบบอ่าน ตามลำดับ และ D คือตัวดำเนินการในการลดอัตราการชักตัวอย่าง B คือตัวดำเนินการในการทำพร่า หรือ การเฉลี่ยภาพ และ E_k คือ *การแปลงระนาบ สัมพรรค* (affine transform) และ $\mathbf{n}_k$ คือสัญญาณรบกวนที่ปรากฏในเฟรมที่ k ตามลำดับ

เราสามารถเขียนสมการที่ 2.1 ได้ใหม่ดังนี้

$$\begin{bmatrix} \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_p \end{bmatrix} = \begin{bmatrix} D_1 B_1 E_1 \\ \vdots \\ D_p B_p E_p \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{n}_1 \\ \vdots \\ \mathbf{n}_p \end{bmatrix} \tag{2}$$

หรือ

$$\begin{aligned} \mathbf{f}_k &= H_k \mathbf{x} + \mathbf{n}_k \\ \mathbf{f} &= \mathbf{H}\mathbf{x} + \mathbf{n}. \end{aligned} \tag{3}$$

2.2.2 อัลกอริธึมส์การสร้างคืน

สมการข้างต้นสามารถแก้ได้ด้วยระเบียบวิธี แก้ปัญหาผกผัน (inverse problem) ที่มีพจน์ *เรกูลาไรเซชัน* (regularization) หรือ

$$\mathbf{x} = \mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{H}^{\mathrm{T}} + \lambda \mathbf{I})^{-1}\mathbf{f}\,. \tag{4}$$

เป็นที่น่าสังเกตว่า เมทริกซ์ **H** เป็น *เมทริกซ์ชนิดมากเลขศูนย์* (sparse matrix) ที่มีขนาดใหญ่มาก ทำให้สมการคณิตศาสตร์ข้างต้นไม่สามารถหาคำตอบได้ง่าย ดังนั้น ระเบียบที่เป็นที่นิยมใช้ในการหาคำตอบนี้ คือ *ระเบียบวิธีความลาดชันสังยุค* (conjugate gradient method)

2.3 การสร้างคืนปริภูมิความละเอียดสูงยวดยิ่ง

การประมวลผลภาพล่วงหน้าส่วนมากที่นิยมใช้ในการรู้จำ หรือ การบีบอัดภาพ คือ การลดขนาดมิติของข้อมูล ในการวิเคราะห์ภาพ PCA เป็นระเบียบวิธีที่เป็นที่นิยมใช้เป็นอย่างมาก ถ้าให้ $\mathbf{\Phi}$ เป็น *ใบหน้าลักษณ์* (eigenface) ที่เล็งเลิศที่ได้ขจัดความซ้ำซ้อนโดยการ *ลดความสัมพันธ์* (decorrelate) ระหว่างข้อมูลภายในภาพ $\mathbf{x}$ ใบหน้าลักษณ์เล็งเลิศ จะถูกเก็บซ่อนไว้เป็นแถวตั้ง ด้วยเหตุนี้ ภาพ $\mathbf{x}$ จึงต้องถูกทำให้เป็นเวกเตอร์ด้วยเช่นกัน ดังนั้น การแทนรูปแบบเล็งเลิศของภาพ $\mathbf{x}$ สามารถเขียนอยู่ในรูป ดังต่อไปนี้

$$\mathbf{x} = \mathbf{\Phi}\mathbf{a} + \mathbf{e}_x\,, \tag{5}$$

โดยที่ $\mathbf{a}$ เป็น ตัวบ่งลักษณ์ที่มีมิติ $L \times 1$ ที่ใช้ในการแทนรูป $\mathbf{x}$ และ $\mathbf{e}_x$ เป็นความผิดพลาดที่เกิดจากการแทนรูป หากกำหนดให้ $\mathbf{\Psi}$ เป็นเมทริกซ์ขนาด $\beta^2 N^2 \times L$ ที่ประกอบด้วย ใบหน้าลักษณ์ของใบหน้าลักษณ์ที่ k^{th} ของเฟรมภาพ LR โดยค่ามาตราส่วนความละเอียด β ซึ่งมีค่าอยู่ระหว่าง

0 ถึง 1 และ N คือจำนวนพิกเซลของภาพใบหน้า เราสามารถเขียนสมการการแทนรูปของภาพความละเอียดต่ำด้วย PCA ได้ดังนี้

$$\mathbf{f}_k = \boldsymbol{\Psi}\hat{\mathbf{a}}_k + \mathbf{e}_{f_k}. \tag{6}$$

โดยการแทนที่ (5) และ (6) ใน (3) เราจะได้

$$\boldsymbol{\Psi}\hat{\mathbf{a}}_k + \mathbf{e}_{f_k} = H_k\boldsymbol{\Phi}\mathbf{a} + H_k\mathbf{e}_x + \mathbf{n}_k. \tag{7}$$

เนื่องจาก

$$\boldsymbol{\Psi}^T\mathbf{e}_{f_k} = 0 \tag{8}$$

และ

$$\boldsymbol{\Psi}^T\boldsymbol{\Psi} = 1 \tag{9}$$

เราสามารถหาสมการดังต่อไปนี้

$$\hat{\mathbf{a}}_k = \boldsymbol{\Psi}^T H_k\boldsymbol{\Phi}\mathbf{a} + \boldsymbol{\Psi}^T H_k\mathbf{e}_x + \boldsymbol{\Psi}^T\mathbf{n}_k. \tag{10}$$

โดยการพิจารณาให้พจน์ที่สองและสามเป็นสัญญาณรบกวนที่ได้จากการสังเกต ให้มีการกระจายแบบเกาส์ (Gunturk et al., 2003) เราสามารถเขียนสมการข้างต้นได้ใหม่ ดังนี้

$$\hat{\mathbf{a}}_k = \Lambda_k\mathbf{a} + \boldsymbol{\Psi}^T\zeta_k \tag{11}$$

โดยที่

$$\zeta_k = H_k\mathbf{e}_x + \mathbf{n}_k \tag{12}$$

และ

$$\Lambda_k = \boldsymbol{\Psi}^T H_k\boldsymbol{\Phi} \tag{13}$$

โดยไม่สูญเสียนัยสำคัญ เราสามารถแก้สมการข้างต้นเพื่อให้ได้ ตัวบ่งลักษณ์ความละเอียดสูงยวดยิ่งที่อยู่ในรูปเวกเตอร์ ในระดับปริภูมิเฉพาะ คล้ายสมการใน หรือ

$$\mathbf{a} = \boldsymbol{\Lambda}^{\mathbf{T}}(\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{\mathbf{T}} + \gamma\mathbf{I})^{-1}\hat{\mathbf{a}} \tag{14}$$

โดยที่ γ เป็นพจน์เรกูราไซชัน เพื่อให้เราสามารถแยกระเบียบวิธีที่จะนำเสนอต่อไป ในที่นี้เราจะใช้ สัญกรณ์เป็นตัวห้อย p ในสมการ (14) เพื่อสื่อความหมายถึง ระเบียบวิธีความละเอียดสูงยวดยิ่งที่ใช้ ในระดับ PCA (Gunturk et al., 2003)

$$\mathbf{a}_p = \boldsymbol{\Lambda}_p^T(\boldsymbol{\Lambda}_p\boldsymbol{\Lambda}_p^T + \gamma\mathbf{I})^{-1}\hat{\mathbf{a}}_p \tag{15}$$

ในบทต่อไป เราจะนำเสนอสัญกรณ์ของระเบียบวิธีที่แตกต่างกันออกไป

2.4 ระบบจำแนกแบบหลายตัว

ในเรื่องที่มีความสำคัญมาก เช่น การเงิน การรักษาพยาบาล สังคม กฎหมาย หรือ เรื่องอื่นๆ ที่มีความซับซ้อน เราจะทำการหาความเห็นที่สอง ก่อนทำการตัดสินใจ บางครั้งเราต้องการความเห็นที่สาม และบางครั้งเราต้องการความเห็นมากกว่านั้น เมื่อเราได้ความเห็นต่างๆมาแล้ว เราจะทำชั่งน้ำหนักความเห็นต่างๆนั้น และ รวบรวมมันเข้าด้วยกันโดยการประมวลความคิดเข้าด้วยกันอย่างใดอย่างหนึ่ง เพื่อให้การตัดสินใจในขั้นสุดท้าย ซึ่งเมื่อผ่านการประมวลผลในขั้นสุดท้ายแล้วน่าจะเป็นการตัดสินใจที่ดีที่สุด (Polikar, 2006) ขบวนการที่ใช้ในการปรึกษา "ผู้เชี่ยวชาญหลายคน" ก่อนการตัดสินใจในขั้นสุดท้าย น่าจะเป็นธรรมชาติที่สองของเราในการดำรงชีวิตในปัจจุบัน จากผลดีที่มีอย่างมากมาย จึงมีการนำมาประยุกต์ใช้กับปัญหาที่ต้องการการตัดสินใจอันโนมัติ ซึ่งได้เกิดเมื่อไม่นานมานี้ในสังคม *ความฉลาดที่คำนวณได้* (computational intelligence)

ระบบจำแนกแบบหลายตัว (multiple classifier systems: MCS) ยังถูกรู้จักกันในชื่อต่างๆ เช่น ระบบเชิงคณะ (ensemble based system) คณะกรรมการตัวจำแนก (committee of classifier) การผสมรวมของผู้เชี่ยวชาญ (mixture of experts) เป็นต้น MCS แสดงให้เห็นผลการทำงานที่ดีกว่า ระบบที่ใช้ตัวตัดสินใจเพียงตัวเดียว ในการประยุกต์ใช้ในด้านต่างๆมากมาย และภายใต้สถานการณ์ที่หลากหลาย

ในบทย่อยนี้ เราจะกล่าวถึง MCS ชนิดต่างๆ เพื่อเป็นการปูพื้นความรู้เกี่ยวกับเรื่องนี้ โดย MCS แต่ละชนิด จะมีความแตกต่างกัน ในเรื่องของอัลกอริธึมส์ ที่ใช้ในการสร้าง ตัวตัดสินใจย่อยแต่ละตัว และ ระเบียบวิธีที่ใช้ในการรวมผลของการตัดสินใจ MCS ที่น่าสนใจ ได้แก่ bagging boosting AdaBoost stacked generalization ระเบียบวิธีปริภูมิย่อยแบบสุ่ม (random subspace method) และ การผสมรวมของผู้เชี่ยวชาญเชิงลำดับชั้น (hiera-rchical) รวมถึงกฎที่ใช้ในการรวมการตัดสินใจ เช่น การลงคะแนนเสียง (Voting) การใช้แม่แบบตัดสินใจ (template decision) หรือ รหัสเอาท์พุตสำหรับการแก้ไขบิตผิดพลาดให้ถูกต้อง (error correcting output codes) เพราะประโยชน์ของ MCS จึงมีการนำไปประยุกต์ใช้ ในการหลอมข้อมูล (data fusion) การเรียนรู้เพิ่มส่วนต่อเนื่อง (incremental learning) รวมถึงหัวข้อวิจัยที่กำลังเป็นที่สนใจในขณะนี้ ของ MCS ที่เกี่ยวข้องกับการคัดเลือกตัวบ่งต่าง และ การเรียนรู้โดยที่มีการขาดหายของตัวบ่งต่าง

ในบทย่อยนี้ เราจะกล่าวถึงปัจจัยสำคัญที่สนับสนุนแนวความคิดของ MCS ทั้งในทางทฤษฎี และทางปฏิบัติ ดังนี้

<u>ทางสถิติ</u> เป็นที่ทราบดีว่า *ข่ายประสาท* (neural networks) หรือ ตัวจำแนกที่ทำงานอัตโนมัติโดยทั่วไปจะประสบปัญหาว่า แม้จะมีสมรรถนะที่ดีเมื่อทำการฝึกฝน แต่สมรรถนะตามนัย

ทั่วไปจะไม่สามารทำนายได้ดีเท่าใดนัก เมื่อทดสอบด้วยข้อมูลที่ตัวจำแนกไม่เคยเห็นมาก่อนในระหว่างการฝึกฝน

ข่ายประสาทแต่ละตัว จะมี นัยทั่วไป ที่แตกต่างกัน หากทำการฝึกฝนด้วยพารามิเตอร์ที่แตกต่างกัน แม้เพียงเล็กน้อย พารามิเตอร์เหล่านี้ ได้แก่ ค่าถ่วงน้ำหนักตั้งต้น หรือ ตัวอย่างฝึกฝนที่แตกต่างกัน หรือ จำนวนพารามิเตอร์ที่ต้องการฝึกฝน (ความซับซ้อน) เป็นต้น ด้วยเหตุนี้ หากเราสามารถเฉลี่ยความสมรรถนะของข่ายประสาทที่ไม่ดีให้ลดน้อยลงได้ เราจะลดความเสี่ยงที่ได้จากการฝึกฝนแล้วได้ตัวจำแนกที่มีประสิทธิภาพต่ำ ถึงแม้ว่า การเฉลี่ยจะไม่สามารถเอาชนะตัวจำแนกที่ดีที่สุด หรือ *ตัวจำแนกที่แข็งแรงที่สุด* (strong classifier) ได้ แต่ที่แน่นอน เราจะสามารถลดความเสี่ยงทั้งหมดที่จะเลือกได้ ตัวจำแนกที่แย่

ข้อมูลขนาดเล็ก ในกรณีที่ข้อมูลฝึกฝนมีจำนวนน้อยมากจนเกินกว่าจะเป็นตัวแทนของการกระจายตัวของข้อมูลที่แท้จริงได้ โดยการใช้เทคนิค *การสุ่มซ้ำ* (resampling) เพื่อให้ได้ เวตย่อยของข้อมูล โดยแต่ละเซตสามารถถูกนำมาใช้ฝึกฝน ตัวจำแนกที่อ่อนแอ (weak classifier) ที่แตกต่างกันจำนวนหนึ่งได้ ตัวจำแนกเหล่านี้จะถูกนำมาใช้ในการตัดสินใจร่วมแบบองค์คณะ แนวความคิดนี้ เสมือนความเห็นของประชาชนผู้น้อยจำนวนมาก ย่อมดีกว่าการตัดสินของเผด็จการ

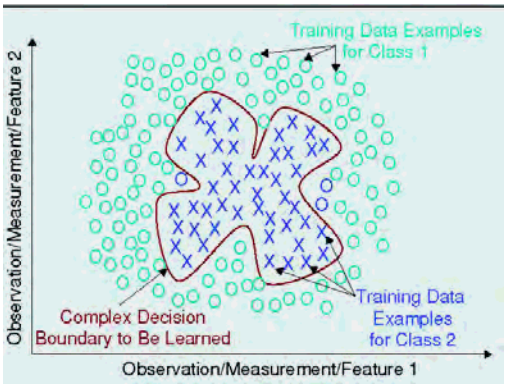
แบ่งแยกและปกครอง ในกรณีที่ ปัญหาที่ต้องการตัดสินใจอาจมีความซับซ้อนจนเกินกว่าตัวจำแนกเพียงตัวเดียวจะสามารถตัดสินใจได้ โดยเฉพาะเมื่อขอบเขตของการตัดสินใจแบ่งคลาสของข้อมูลมีความซับซ้อน หรือ อยู่นอกเหนือขอบเขตของฟังก์ชันที่จะใช้ในการเลือกแบบจำลองตัวจำแนก เช่น ตัวจำแนกเราเป็นชนิดเชิงเส้น แต่ *ขอบเขตการตัดสินใจ* (decision boundary) ของเราจะแบบไม่เชิงเส้นที่มีความซับซ้อน ดังแสดงในรูปที่ 2.10

เนื่องจากขอบเขตการตัดสินใจมีความซับซ้อนมาก หากเราสามารถเลือกได้ ตัวจำแนกที่แข็งแรงขึ้นกว่าตัวจำแนกเชิงเส้น ซึ่งในที่นี้หมายถึง ตัวจำแนกที่มีรูปร่างของขอบเขตการตัดสินใจเป็นแบบ วงรี หรือ วงกลม จะเห็นได้ว่า ตัวจำแนกอ่อนแอ ที่สร้างขึ้นแต่ละตัวจะขึ้นกับขอบเขตของข้อมูลที่ใช้ในการฝึกฝน เมื่อนำขอบเขตการตัดสินใจของตัวจำแนกทั้งหมด มาร่วมกัน เราจะได้ขอบเขตการตัดสินใจในขั้นสุดท้าย ที่สามารถประมาณขอบเขตการตัดสินใจที่แท้จริงได้ ดังแสดงในรูปที่ 2.11

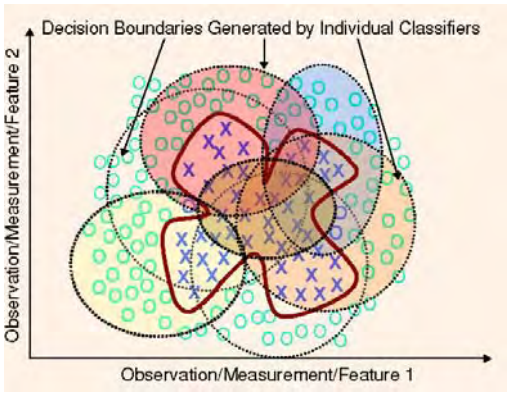
2.4.1 ความหลากหลาย

ความหลากหลาย (diversity) เป็นองค์ประกอบที่สำคัญจำเป็นสำหรับการรู้จำด้วย MCS การตัดสินใจที่ผิดพลาดของตัวจำแนกแต่ละตัวของ MCS ควรจะต้องแตกต่างกัน โดยธรรมชาติความแตกต่างของความผิดพลาดของการตัดสินใจของ ตัวจำแนกแต่ละตัวนั้นเมื่อเกิดขึ้น และถูกรวมเข้าด้วยกันอย่างมีระบบ เราจะสามารถลดความผิดพลาดในการตัดสินใจรวมทั้งหมดลงได้

แนวความคิดนี้เป็นแนวความคิดที่ค่อนข้างจะเหมือนกับการกรองผ่านต่ำ (low pass)



รูปที่ 2.10 ขอบเขตการตัดสินใจที่ซับซ้อนที่ตัว จำแนกเชิงเส้นไม่สามารถแบ่งแยกได้



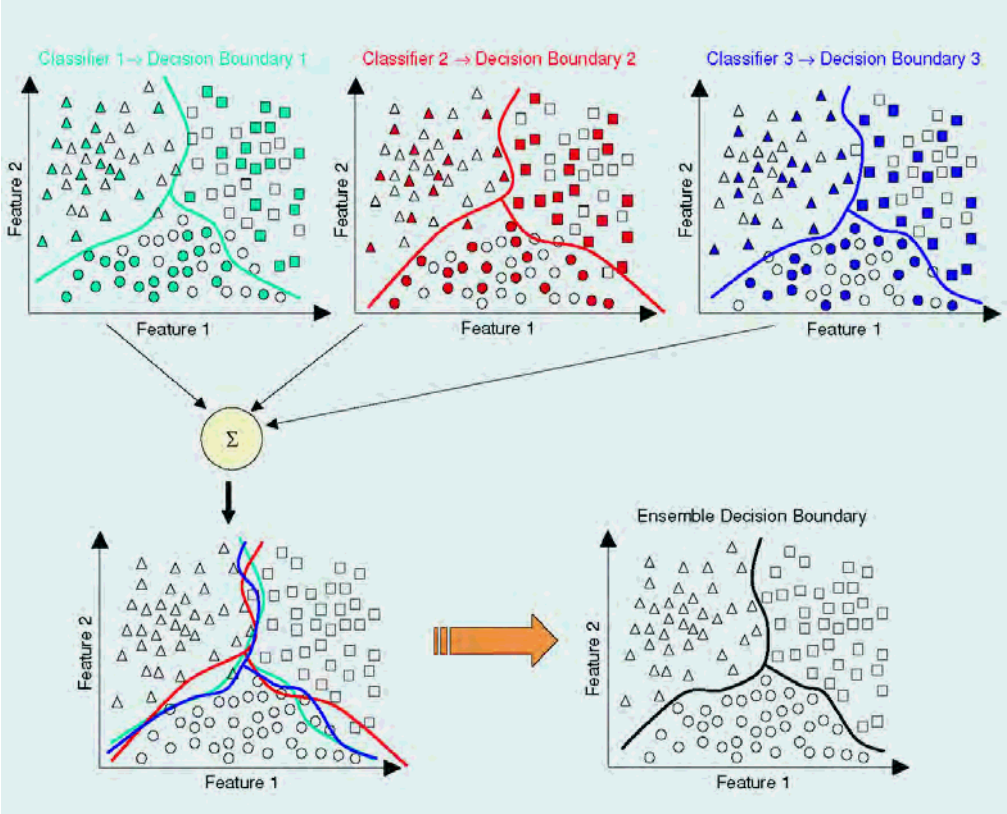
รูปที่ 2.11 คณะกรรมการของตัวจำแนกที่แผ่ทั่วปริภูมิของการตัดสินใ

สัญญาณออกจากสัญญาณรบกวน โดยสรุป MCS ต้องการตัวจำแนกที่ ขอบเขตการตัดสินใจแต่ละตัว แตกต่างจากขอบเขตการตัดสินใจของตัวจำแนกตัวอื่น ชุดของตัวจำแนกนี้จึงจะถูกพิจารณาว่ามี *ความหลากหลาย*

ดังแสดงในรูปที่ 2.12 เราได้สร้างตัวจำแนกขึ้นมา 3 ตัว ตัวจำแนกแต่ละตัวมีขอบเขตการตัดสินใจที่แตกต่างกัน ซึ่งเกิดจากข้อมูลการฝึกฝนที่แตกต่างกัน แม้ทว่าตัวจำแนกแต่ละตัวจะสามารถสร้างขอบเขตการตัดสินใจที่มีสมรรถนะสูง สามารถแบ่งคลาสของข้อมูลที่ใช้ฝึกฝนได้อย่างถูกต้องแม่นยำ แต่หากเราเลือกใช้ตัวจำแนกตัวใดตัวหนึ่งจากตัวจำแนกทั้งามตัวนี้ในการทดสอบตัวอย่างซึ่ง ตัวจำแนกนั้น ไม่เคยเห็นมาก่อน จะเห็นว่า เรามีความผิดพลาดในการทำนายเท่ากับ 4 สำหรับตัวจำแนกตัวที่ 1 (1 สามเหลี่ยม และ 3 วงกลม) และ ความผิดพลาดเท่ากับ 4 สำหรับตัวจำแนกตัวที่ 2 (1 สามเหลี่ยม 2 วงกลม และ 1 สี่เหลี่ยม) เป็นต้น แต่เมื่อเราตัวจำแนกทั้งสามตัวมารวมกัน เราจะได้ ผลรวมของขอบเขตการตัดสินใจ ที่สามารถแบ่งคลาสดดยไม่มีความผิดพลาดเกิดขึ้นเลย

2.5 การคัดเลือกลักษณะบ่งต่าง

การสกัดตัวบ่งต่าง (feature extraction) *และ การคัดเลือกตัวบ่งต่าง* (feature selection) เป็นสองส่วนสำคัญที่ทำหน้าที่ต่างกันในการรู้จำรูปแบบ โดยทั่วไป การสกัดตัวบ่งต่างเน้นไปที่การแปลงตัวบ่งต่างที่มีอยู่ ให้เป็นตัวบ่งต่างชุดใหม่ที่แต่ละตัวต่างก็มีข่าวสารที่สำคัญและไม่ซ้ำซ้อนกัน จากนั้นจะเลือกเฉพาะตัวบ่งต่างที่มี กำลังงานรวมสูง หรือ คุณสมบัติในการรู้จำสูง เนื่องจากจำนวนตัวบ่งต่างใหม่ที่ถูกเลือกมีจำนวนน้อยลง ระเบียบวิธีการสกัดตัวบ่งต่างจึงทำการลดขนาดมิติจากขนาดมิติของตัวบ่งต่างเดิมที่มีขนาดใหญ่ ไปเป็นตัวบ่งต่างใหม่ที่มีขนาดมิติเล็กลง การลดขนาดมิติจะมีสมรรถนะที่ดีเมื่อใช้กับตัวบ่งต่างที่มีความสัมพันธ์ที่ซ้ำซ้อนกัน



รูปที่ 2.12 การรวมตัวจำแนกที่เรียนรู้ด้วยข้อมูลเซตย่อยของข้อมูลฝึกฝนที่แตกต่างกัน

มองอีกมุมหนึ่ง การสกัดตัวบ่งต่างจะให้ผลดีในกรณีที่ชุดตัวบ่งต่างที่พิจารณามีสหสัมพันธ์ต่อกันค่อนข้างสูง ในขณะที่ การคัดเลือกตัวบ่งต่าง หรือ การคัดเลือกตัวแปร จะมีสมรรถนะที่ไม่น่าพอใจนัก เมื่อใช้กับชุดของตัวบ่งต่างในลักษณะดังกล่าว ทั้งนี้เป็นเพราะว่า ในการคัดเลือกตัวบ่งต่าง ตัวบ่งต่างที่มีสหสัมพันธ์ที่ใกล้กันจะได้ *แต้ม* (score) ที่ใกล้กัน หากแต่ว่าเมื่อตัวบ่งต่างทั้งสองมีค่าของแต้มสูง ตัวบ่งต่างทั้งสองก็จะถูกเลือก ซึ่งในความเป็นจริง ตัวบ่งต่างทั้งสองมีปริมาณข่าวสารที่สำคัญต่อการรู้จำที่ละม้ายคล้ายกัน เพราะมีสหสัมพันธ์ที่ใกล้เคียงกัน ซึ่งในกรณีของการรู้จำในกรณีนี้นั้น ตัวบ่งต่างตัวหนึ่งแค่หนึ่งตัวก็เพียงพอต่อการรู้จำแล้ว เพราะมีปริมาณข่าวสารที่สำคัญเกือบครบถ้วน หากเพิ่มตัวบ่งต่างที่มีแต้มเท่ากันที่เหลือ ก็จะเป็นการเพิ่มข่าวสารที่ซ้ำซ้อนซึ่งจะไม่เป็นประโยชน์เท่าใดนักในการรู้จำ [21]

จากความรู้พื้นฐานในบทย่อยที่ผ่านมา เราได้ทราบถึงระเบียบวิธีในการลดขนาดมิติตัวบ่งต่างมาแล้ว สำหรับการคัดเลือกตัวบ่งต่างนั้น สามารถแยกออกได้เป็นสองแนวทาง (Wolf, 2005 & Guyon, 2003) แนวทางแรกเรียกว่า แบบฟิวเตอร์ (filter method) ในแนวทางนี้ตัวบ่งต่างจะถูก *จัดลำดับความสำคัญ* (ranking) ด้วย *แต้ม* (score) จากมากไปหาน้อย แต้ม หรือ คะแนน นี้ จะถูกคำนวณด้วยมาตรที่กำหนดขึ้น โดยเช่น อาจเป็นค่า *สัมประสิทธิ์สหสัมพันธ์* (correlation coefficient) หรือ *ข่าวสารร่วมกัน* (mutual information) เป็นต้น

แนวทางที่สอง เรียกว่า แบบภาชนะ (wrapper method) ที่ทำการคัดเลือกชุดย่อยของตัวบ่งต่าง โดยการทดสอบผลตอบรับที่ดีจาก ตัวทำนาย (ตัวจำแนก) ต่อตัวบ่งต่างที่เลือก

สำหรับการคัดเลือกตัวบ่งต่าง ในแบบฟิวเตอร์นั้น การจัดลำดับความสำคัญของตัวบ่งต่างนั้นเป็นการประมวลผลขั้นต้น ที่ไม่จำเป็นต้องขึ้นอยู่กับตัวทำนาย การคัดเลือกตัวบ่งต่างในลักษณะนี้จึงมีประสิทธิภาพในการคำนวณค่อนข้างมาก และ ดีที่สุดต่อตัวทำนายใดๆ โดยนัยทั่วไป (generalization) ทั้งนี้เนื่องมาจากคุณสมบัติเชิงตั้งฉาก และ ความเป็นอิสระต่อกัน ส่วนการคัดเลือกตัวบ่งต่างแบบภาชนะ จำเป็นต้องขึ้นอยู่กับตัวทำนาย และต้องมีข้อมูลชุดพิเศษเพื่อไว้ใช้ในการทดสอบผลตอบรับ ซึ่งไม่ค่อยจะเหมาะกับการรู้จำใบหน้าเท่าใดนัก เนื่องจากจำนวนข้อมูลของภาพใบหน้าของเรามีจำนวนน้อยเมื่อเทียบกับขนาดของมิติ

ในการรู้จำใบหน้า เราจะมีตัวอย่างเพียงสองสามร้อยตัวอย่างสำหรับใช้ทั้งฝึกฝนและทดสอบไปพร้อมกัน แต่ขนาดของมิติของภาพจะมีขนาดตั้งแต่สองสามพัน ไปจนถึง หลายแสน ถ้าเราใช้ฟิลเตอร์ หรือ การประมวลผลทางภาพ เช่น ผลการแปลงเวฟเลต มาใช้เราอาจลดขนาดของมิติลงไปเหลือเพียงไม่กี่พัน นอกจากการคัดเลือกตัวบ่งต่างจากตัวบ่งต่างของภาพใบหน้าแล้ว ในกรณีที่เราสามารถสร้างชุดของตัวบ่งต่างที่แปรเปลี่ยนตามทิศ (ระบุทิศทาง) เราสามารถใช้ระเบียบวิธีในการคัดเลือกตัวแปร ในการคัดเลือกชุดของตัวบ่งต่างแทนได้ หรือ อีกนัยหนึ่ง คือ เราเลื่อนการคัดเลือกตัวบ่งต่างขึ้นไปอีกหนึ่งระดับ หมายถึงเราเลื่อนจากการคัดเลือกตัวบ่งต่างไปเป็นการคัดเลือกชุดของตัวบ่งต่างแทน

เนื่องด้วย การคัดเลือกตัวบ่งต่างแบบฟิวเตอร์ ค่อนข้างมีความเหมาะสมกับการรู้จำวัตถุกว่าแบบภาชนะ หนึ่งในเงื่อนไขที่สำคัญในการจัดลำดับความสำคัญของแบบฟิวเตอร์ ได้แก่ *สัมประสิทธิ์สหสัมพันธ์เพียรสัน* (Pearson correlation coefficient)

$$R(i) = \frac{COV(X, Y_i)}{\sqrt{\mathrm{var}(X)\,\mathrm{var}(Y_i)}} \tag{16}$$

เงื่อนไขที่สำคัญในการจัดลำดับความสำคัญอีกตัวหนึ่งที่เกี่ยวกับ *ทฤษฎีข่าวสาร* (information theory) ซึ่งเกี่ยวข้องกับการประมาณเชิงประจักษ์ของ ข่าวสารร่วมกัน

$$I(i) = \int_{x_i}\int_{y} p(x_i, y) \log \frac{p(x_i, y)}{p(x_i)p(y)} dxdy\,, \tag{17}$$

โดยที่ $P(x_i)$ และ $P(y)$ เป็นความน่าจะเป็นของ x_i และ y และ $P(x_i,y)$ เป็นฟังก์ชันความหนาแน่นความน่าจะเป็นร่วม (joint probability density function) เงื่อนไขนี้เกี่ยวข้องกับ *เอนโทปี* (entropy) ของหลักการวิศวกรรมไฟฟ้าสื่อสาร

เงื่อนไขข้างต้นเป็นมาตรวัดความเป็นอิสระต่อกันระหว่าง x_i ความหนาแน่นของตัวแปรสุ่ม กับ ความหนาแน่นตัวแปรสุ่มเป้าหมาย *y* เนื่องจาก $P(x_i)$ $P(y)$ และ $P(x_i,y)$ ไม่สามารถทราบค่าที่แท้จริงได้ เราจึงต้องคำนวณความหนาแน่นความน่าจะเป็นดังกล่าวโดยการประมาณ

จากข้อมูล ในกรณีนี้เราจะได้ ข่าวสารร่วมกัน จากการคำนวณเชิงประจักษ์ โดยการแทนที่อินทิกรัล ด้วย ผลรวม ดังนี้

$$I(i) = \sum_{x_i} \sum_{y} P(X = x_i, Y = y) \log \frac{P(X = x_i, Y = y)}{P(X = x_i) P(Y = y)} \tag{17}$$

อย่างไรก็ดี เงื่อนไขข้างต้น เป็นเงื่อนไขพื้นฐาน ยังมีเงื่อนไข นอกเหนือจากนี้ที่เป็นเงื่อนไขที่ดีต่อการรู้จำ เช่น เอนโทปี ที่สามารถวัดความแตกต่างของความหนาแน่นความน่าจะเป็นระหว่างเวกเตอร์ตัวแปรสุ่มสองตัว เอนโทปี ที่ว่านี้มีคุณสมบัติใน *การแยกแยะ* (discriminant) ทางสถิติระหว่างคลาสที่ดี เป็น มาตรวัด *เอนโทรปีสัมพัทธ์* (relative entropy) ที่วัดความแตกต่างของการกระจายตัวแปรสุ่มที่สนใจ เอนโทปี ที่ว่านี้ รู้จักกันดีว่า เอนโทปีไขว้ (cross entropy) หรือ *ระยะห่าง Kullback-Leibler* (Kullback-Leibler distance) หรือ I-divergence (Saito, 1995 และดูเอกสารอ้างอิงในบทความฉบับนี้)

$$I(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \tag{18}$$

โดยที่ $\mathbf{p} = \{p_i\}_{i=1}^{n}$ และ $\mathbf{q} = \{q_i\}_{i=1}^{n}$ เป็นลำดับที่มีค่าไม่เป็นศูนย์ (อาจอยู่ในรูปของกำลังงาน) และ $\sum p_i = \sum q_i = 1$ (พิจารณาเหมือนกับการนอมอลไรซ์ให้การกระจายตัวของกำลังงานของสัญญาณของตัวบ่งต่างที่ 1 และ 2 ให้มีค่าเท่ากัน) ในกรณีที่เราจะประยุกต์ใช้มาตรวัดนี้ และ เพื่อความรวดเร็วในการคำนวณ เราจะกำหนดให้ ค่ามาตรวัดระหว่างตัวบ่งต่างที่คำนวณได้จากสมการที่ (18) มี *คุณสมบัติเชิงบวก* (additive) สมการที่ (18) สำหรับคำนวณการจัดลำดับความสำคัญ ที่ปรับปรุงใหม่จะเป็นดังนี้

$$\tilde{I}(\mathbf{p}, \mathbf{q}_1^k) = \sum_{j=1}^{k} \sum_{i=1}^{n} p_i \log \frac{p_i}{q_{i,j}} \tag{19}$$

โดยที่ *k* คือ จำนวนตัวบ่งต่าง (ชุดของตัวบ่งต่าง) ในขณะของ ระบบจำแนกแบบหลายตัว

ข้อน่าสังเกต เอนโทรปี ที่มีในสมการที่ (18) มีความไม่สมมาตรระหว่าง $\mathbf{p}$ และ $\mathbf{q}$ อยู่ ในบางกรณี เราควรใช้ เอนโทรปี ที่มีความสมมาตร หรือ ใช้ J-divergence ระหว่าง $\mathbf{p}$ และ $\mathbf{q}$ ดังนี้

$$\tilde{J}(\mathbf{p}, \mathbf{q}) = \tilde{I}(\mathbf{p}, \mathbf{q}_1^k) + \tilde{I}(\mathbf{q}_1^k, \mathbf{p}) \tag{20}$$

หรือแม้แต่ อย่างง่าย เช่น

$$W(\mathbf{p}, \mathbf{q}) = \|\mathbf{p}, \mathbf{q}\|^2 \tag{21}$$

ก็สามารถนำมาใช้แทนสมการที่ (18) เพื่อใช้ในการคัดเลือกตัวบ่งต่างได้

อย่างไรก็ดี มาตรวัดที่กล่าวมาข้างต้น ถ้าไม่เน้นการวัดคุณสมบัติทางด้านการแยกแยะ ก็เน้นการวัดความไม่เป็นอิสระต่อกัน เป็นที่ทราบดีว่า ในระบบจำแนกแบบหลายตัว การรวมการตัดสินใจที่มี *สหสัมพันธ์เชิงลบ* (negative correlation) จะให้ผลการตัดสินใจสุดท้ายที่ดีกว่าเงื่อนไขข้างต้น ดังนั้นมาตรวัดที่ดีที่สุดในการตัดสินใจน่าจะเป็นมาตรวัดที่สามารถวัดการแยกแยะที่มีข่าวสารเกี่ยวกับ สหสัมพันธ์เชิงลบ อยู่ด้วย ซึ่งเป็นส่วนหนึ่งของงานวิจัยที่ดำเนินการอยู่

การคัดเลือกตัวบ่งต่าง เดิมเป็นแนวความคิดในการคัดเลือกตัวบ่งต่าง ส่วนมาตรวัดความสัมพันธ์เชิงลบ เป็นมาตรวัด ความหลากหลาย ที่ใช้ในการคัดเลือกตัวจำแนกสำหรับ MCS หากเราพิจารณาให้ ตัวจำแนกแต่ละตัวเสมือนเป็น ลักษณะบ่งต่างเชิงลำดับชั้น ในลำดับชั้นที่สูงกว่า ตัวบ่งต่างที่สกัดได้จาก 2DPCA เราจะอนุโลมให้การคัดเลือกตัวจำแนกสำหรับ MCS สามารถดำเนินการด้วยระเบียบวิธี ของการคัดแยกตัวบ่งต่าง

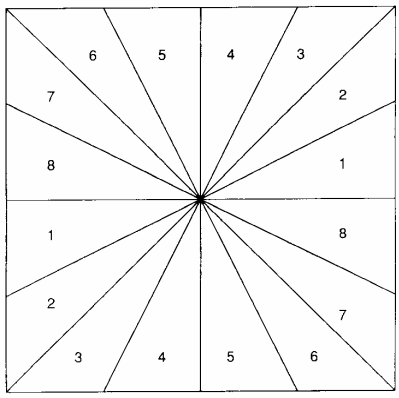
2.6 ผลการแปลงเวฟเลตระบุทิศทาง

ความสำเร็จของผลการแปลงเวฟเลตนั้น เนื่องมาจากความสามารถในการแทนรูปฟังก์ชันที่ราบเรียบเป็นช่วง (piecewise smooth function) โดยที่ใช้จำนวนสัมประสิทธิ์จำนวนไม่มาก ทำให้เวฟเลตเป็นเครื่องมือชนิดใหม่ที่นิยมใช้ในการประมวลผลสัญญาณและภาพ ผลการแปลงเวฟเลตสองมิตินั้นสามารถคำนวณได้อย่างรวดเร็วด้วย *ผลเทนเซอร์* (tensor product) ของผลการแปลงเวฟเลตหนึ่งมิติ ด้วยเหตุนี้ *เวฟเลตแบบแยกออกจากกันได้* (separable wavelet) จึงเป็นผลการแปลงเวฟเลตที่นิยมในปัจจุบัน

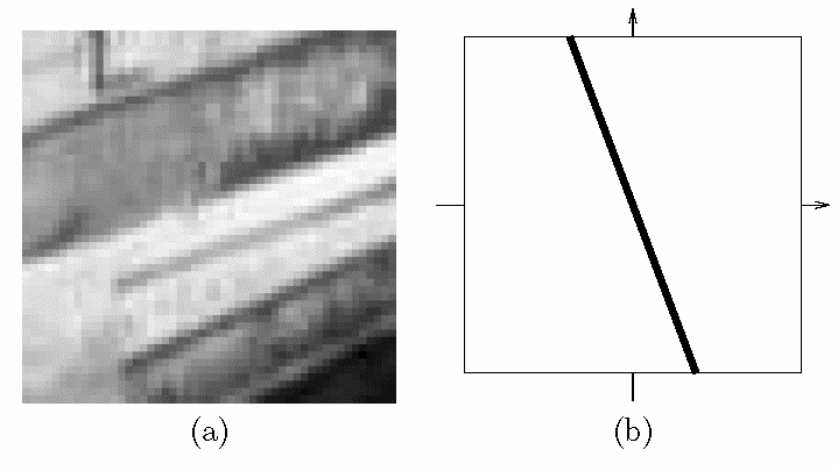
แนวความคิดของผลการแปลงเวฟเลตระบุทิศทาง ได้มาจาก ฟิวเตอร์พัด (Fan filter) ที่ประสบความสำเร็จในการประยุกต์ใช้ทางหุ่นยนต์ และ การมองเห็นด้วยคอมพิวเตอร์ การวิเคราะห์ทางธรณีวิทยา และ ตัวจับลักษณะบ่งต่างเชิงเส้นและการปรับปรุงคุณภาพสัญญาณ รวมถึงการใช้งานในทางบีบอัดภาพ แนวความคิดฟิวเตอร์พัด ได้ถูกปรับปรุงให้ดำรงคุณสมบัติของผลการแปลงเวฟเลตโดย Bamberger (Bamberger, 1992) รูปที่ 2.13 แสดงฟิวเตอร์พัดที่ดำรงคุณสมบัติของผลการแปลงเวฟเลต จากนั้น ผลการแปลงเวฟเลตแบบระบุทิศทางแบบต่างๆก็ได้พัฒนาต่อเนื่อง เช่น Bandelet pyramid directional wavelets และ Curvelet เป็นต้น

อย่างไรก็ดี ผลการแปลงเวฟเลตแบบแยกออกจากกันได้ ที่กล่าวข้างต้นนี้ ยังไม่สามารถให้การแทนรูปอย่างมากเลขศูนย์ (sparse representation) กับภาพที่มีโครงสร้างต่างกันออกไปได้ แม้ทว่าผลการแปลงเวฟเลต จะสามารถให้การแทนรูปที่มากเลขศูนย์ ต่อ *ภาวะเอกฐาน* (singularity) ในทิศแนวระนาบราบและดิ่ง ได้ (จากการใช้ เวฟเลตในแนวระนาบราบและดิ่ง) แต่มันยังไม่สามารถแทนรูปได้อย่างสมบูรณ์ต่อ ภาวะเอกฐานที่คงอยู่ในทิศทางอื่น ด้วยเหตุนี้ จึงมีการศึกษาผลการแปลงเวฟเลตแบบมีทิศทาง รูปที่ 2.14 แสดงผลตอบสนองทางความถี่

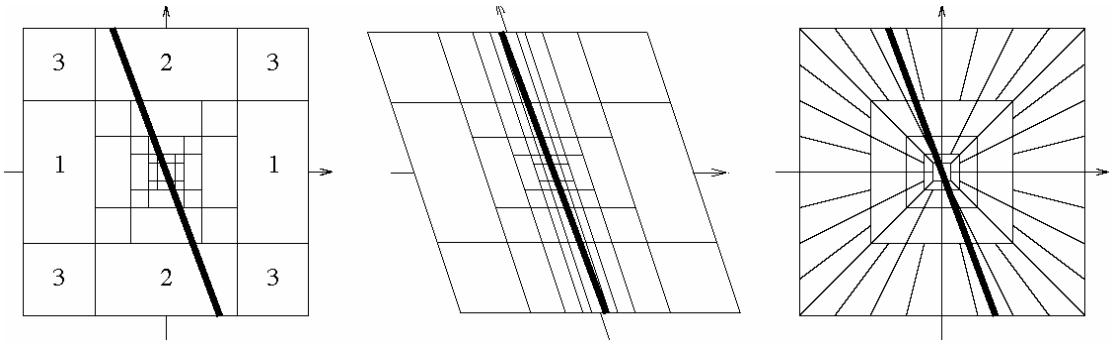
ของภาพในรายละเอียด ภาพทางซ้ายเป็นภาพในแนวปริภูมิตำแหน่ง ส่วนภาพทางขวาเป็นผลตอบสนองทางความถี่ ตามที่แสดงในภาพ จะเห็นว่าผลตอบสนองทางความถี่มีทิศทางที่แน่นอน หรือ อาจถูกเรียกว่า *ทิศของการไหลของโครงสร้างของภาพ* (geometrical flow)



รูปที่ 2.13 แถบความถี่ย่อย 8 แถบที่ถูกแบ่งแบบมีทิศทาง



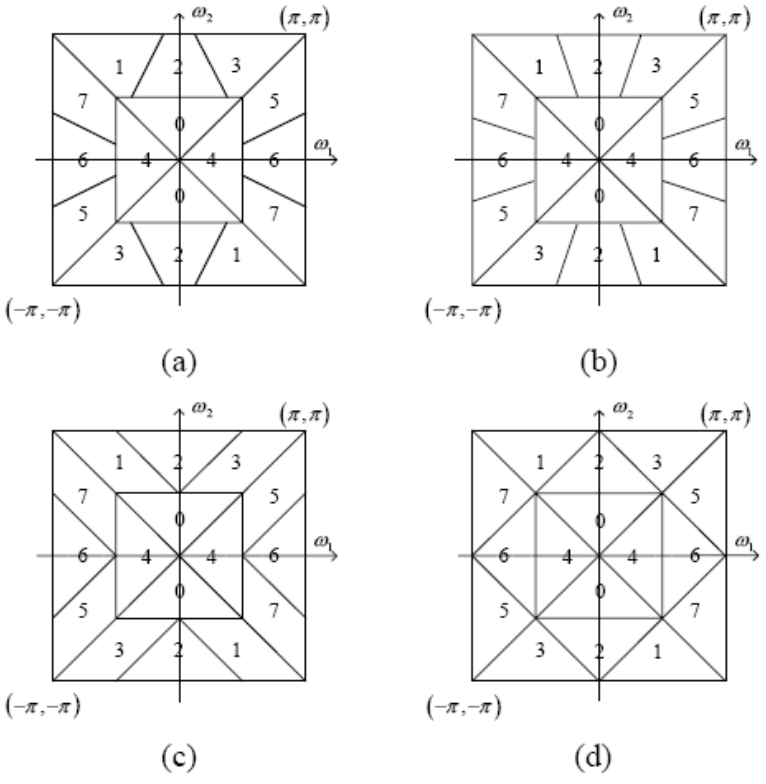
รูปที่ 2.14 รายละเอียดภาพและสเปกตรัมที่หนาแน่นเป็นแนวเส้นตรง



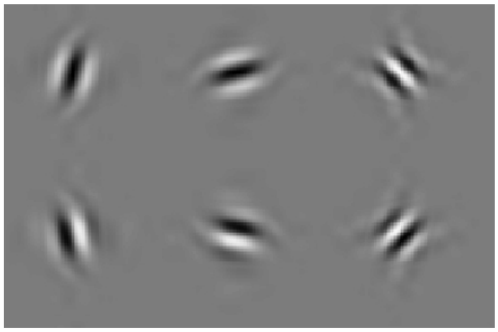
รูปที่ 2.15 ผลตอบสนองทางความถี่ จากซ้ายไปขวา ผลการแปลงเวฟเลต bandelets (รุ่นที่ 1) และ Curvelet

จากการสังเกตภาพโดยทั่วไปพบว่า ทิศของการไหลของโครงสร้างของภาพ มีการกระจายตัวไปทั่วทิศทางอย่างเป็นเอกรูป (uniform) โดยไม่ขึ้นกับ สเกล (scale) หรือ การซูมเข้า (zoom in) ซึ่งเป็นพารามิเตอร์ที่สำคัญของผลการแปลงเวฟเลต รูปที่ 2.15 แสดงให้เห็นว่า เมื่อใช้ผลการแปลงเวฟเลต (รูปซ้ายสุด) กำลังงานที่ได้จะกระจายไปทั่ว จึงมีสัมประสิทธิ์จำนวนมากที่มีค่าไม่เป็นศูนย์ รูปกลาง แสดง Brandelets ที่เป็นผลการแปลงเวฟเลตที่ปรับตัวได้ ซึ่งจะปรับตัวตาม geometrical flow ของภาพ ส่วนรูปขวาสุด เป็น Pyramid directional wavelet ที่พัฒนาจากฟิวเตอร์พัดของ Bamberger

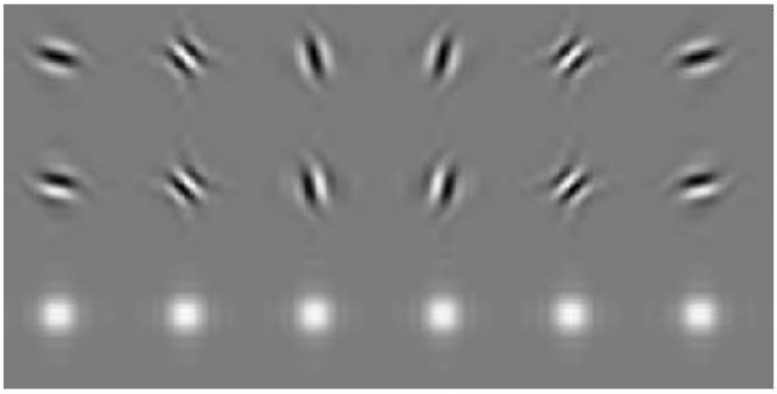
ในปัจจุบันได้มีการพัฒนา ผลการแปลงเวฟเลตระบุทิศทาง (Nguyen, 2007 & Durand 2008) อีกหลายแบบ รูปที่ 2.16 เป็นหนึ่งในความเป็นไปได้ในการสร้างผลการแปลงเวฟเลตระบุทิศทาง หรือ เวฟเลตสองมิติ ชนิด dual-tree complex (Selesnick, n.d.) ซึ่งได้มีการพัฒนาทั้งในแบบชนิดจำนวนจริง และ จำนวนเชิงซ้อน เวฟเลตสองมิติชนิดจำนวนจริง จะมีการสร้างเวฟเลตที่ขึ้นกับทิศ รูปที่ 2.17 แสดงเวฟเลตจำนวนจริงจำนวน 6 ทิศ ทิศละ 30 องศา ส่วนเวฟเลตชนิดจำนวนเชิงซ้อน ในแต่ละทิศจะประกอบด้วย เวฟเลต จำนวน 2 ตัว เวฟเลตตัวแรกสำหรับจำนวนจริง และตัวที่สองสำหรับจำนวนเชิงซ้อน เมื่อประกอบกันเราจะได้เวฟเลตที่ขึ้นกับทิศทาง รูปที่ 2.18 แสดงเวฟเลตจำนวนเชิงซ้อน แถวนอนที่สามแสดงขนาด (magnitude) ผลการแปลงเวฟเลตทั้งสองให้ผลดีสามารถลด สิ่งแปลกปน (artifact) แบบตารางหมากรุกได้เป็นอย่างดี



รูปที่ 2.16 ผลตอบสนองทางความถี่ 4 แบบ ของผลการแปลงเวฟเลตระบุทิศทาง



รูปที่ 2.17 เวฟเลตระบุทิศทาง ชนิดจำนวนจริง (Selesnick, n.d.)



รูปที่ 2.18 เวฟเลตระบุทิศทาง ชนิดจำนวนเชิงซ้อน (Selesnick, n.d.)

การใช้ คณะกรรมการตัดสินใจ ร่วมกับผลการแปลงเวฟเลต (Asdornwised,2002) ได้ถูกนำเสนอเป็นครั้งแรก โดยใช้การลงคะแนนเสียง ไม่มีการคัดเลือกตัวบ่งต่างที่อยู่ในแต่ละแถบความถี่ย่อย อย่างไรก็ดี ได้มีการค้นพบว่า ความสามารถในการแยกแยะวัตถุ ของแต่ละแถบความถี่ย่อยมีอำนาจหรือคุณภาพที่ไม่เท่ากัน และ ความสามารถในการแยกแยะสูงสุดไม่จำเป็นต้องอยู่ที่แถบความถี่ย่อยที่มีผลตอบสนองทางความถี่ต่ำที่สุด ในเอกสารอ้างอิงในบทความที่วิจัย (Sermwuthisarn, 2007) ได้ทำการทดลองการควบรวมการตัดสินใจ โดยทำการเลือกอย่างไม่ฉลาด (exhaustive search) ในทุกการ *จัดหมู่* (combinations) ของแถบความถี่ย่อย

คณะผู้วิจัย (Sermwuthisarn, 2007 & Sermwuthisarn, 2008) ได้นำเสนอวิธีคัดเลือกแถบความถี่ย่อยโดยอาศัยเงื่อนไขความเป็นอิสระต่อกัน ในการเลือกแถบความถี่ย่อยสำหรับการรู้จำใบหน้า อนึ่งเนื่องจาก ในงานวิจัยเหล่านี้ใช้ผลการแปลงเวฟเลตที่ไม่ระบุทิศทาง หากเราใช้ผลการแปลงเวฟเลตแบบระบุทิศ เราน่าจะได้สมรรถนะในการรู้จำที่ดีขึ้น

3. การสร้างคืนปริภูมิย่อยสองมิติความละเอียดสูงยวดยิ่งสำหรับการรู้จำวัตถุ

ในบทนี้ จะกล่าวถึง การสร้างคืนความละเอียดสูงยวดยิ่งในระดับปริภูมิย่อย โดยมีการพัฒนาเพิ่มเติมด้วย 2DPCA แบบระบุทิศทาง หรือ 2DPCA แบบไขว้ที่สามารถเก็บรายละเอียดได้หลากหลายขึ้น เพื่อประกอบกันขึ้นเป็นคณะกรรมการในการตัดสินใจ โดยในขั้นแรกเราจะทำความเข้าใจความแปรปรวนร่วมเกี่ยวไขว้ จากนั้นจะนำเสนอระเบียบวิธีในการสร้างคืนปริภูมิความละเอียดสูงยวดยิ่งสองมิติ ซึ่งจำเป็นต้องใช้ความรู้พื้นฐานเกี่ยวกับตัวดำเนินการทำพร่าเกาส์ และ แบบจำลองทางคณิตศาสตร์ที่เหมาะสม

ขั้นที่สอง เราจะแนะนำการรู้จำด้วย MCS แบบระเบียบวิธีปริภูมิย่อยแบบสุ่ม (random subspace method: RSM) และใช้การรวมการตัดสินใจแบบลงคะแนนเสียง เนื่องจากเราใช้วิธีในการเลือกปริภูมิย่อยแบบสุ่มซึ่งได้ก่อให้เกิด *ความหลากหลาย* (diversity) ในการตัดสินใจอยู่แล้ว เนื่องด้วยคุณสมบัติความหลากหลายในการตัดสินใจที่เกิดขึ้นมีการกระจายตัวที่เป็นอิสระต่อกัน เราจึงสามารถรวมการตัดสินใจได้ด้วยการลงคะแนนเสียง เชิงประชาธิปไตย โดยไม่ต้องมีการถ่วงน้ำหนักไปที่การตัดสินใจของตัวจำแนกตัวหนึ่งตัวใดเป็นพิเศษ หรือ จำเป็นต้องทำการคัดเลือกชุดตัวบ่งต่าง

สุดท้าย เราจะไม่ทำการเลือกปริภูมิย่อย ด้วย RSM แต่เราจะคัดเลือกชุดของตัวบ่งต่างจำนวนหนึ่งที่เข้าใจว่าได้เก็บรายละเอียดที่สำคัญและไม่ซ้ำซ้อนจนเกินไป มาทำการสร้างการตัดสินใจแบบองค์คณะ เราจึงต้องมีการคัดเลือกตัวบ่งต่าง เพื่อเพิ่มประสิทธิ์ภาพในการรู้จำภาพวัตถุ ข้อควรสังเกตุ คือระเบียบวิธีที่นำเสนอทั้งสองนี้จะกระทำในระดับปริภูมิย่อย 2DPCA และ 2DPCA รุ่นถัดมา เป็นหลัก

3.1 การสร้างเมทริกซ์ความแปรปรวนไขว้

ไม่เป็นการยากที่จะทำความเข้าใจว่า PCA ทำการเข้ารหัสข่าวสารโดยการเข้าคู่ความสัมพันธ์ระหว่างคู่พิกเซล (ดูรูปที่ 3.1 ที่แสดงต่อไปเพื่อประกอบความเข้าใจ) ในขณะที่ 2DPCA และระเบียบวิธีในรุ่นถัดไป ตั้งใจเข้ารหัสแบบจำลองโดยการเลือกปริภูมิตำแหน่งเฉพาะของภาพใบหน้าในบางปริภูมิตำแหน่ง (ทิศ) การเข้ารหัสข่าวสารในลักษณะนี้จะคงทนกว่าปริภูมิตำแหน่งที่ไม่ได้เลือก เป็นเพราะว่า ข่าวสารในปริภูมิตำแหน่งที่ไม่ถูกเลือกจะถูกรวบรวมสะสมเป็นหนึ่ง คงเหลือแต่ลักษณะบ่งต่างในทิศที่มีความสำคัญต่อการรู้จำที่ถูกเข้ารหัส ยกตัวอย่างเช่น จากความรู้เกี่ยวกับความเป็นไปได้ทางชีวภาพ ของ 2DPCA แบบระบุทิศทาง

2DPCA ที่เน้นแบบจำลองทางแถวตั้ง (column) ควรจะสามารถเข้ารหัสที่สำคัญต่อการรู้จำได้ดีกว่า 2DPCA ที่เน้นแบบจำลองทางแถวนอน (row) ซึ่งเมื่อเราขยายการสกัดตัวบ่งต่างให้ระบุทิศได้มากขึ้นกว่า แนวระนาบ และ แนวดิ่ง ให้มีทิศมากขึ้น ตัวบ่งต่างที่สกัดได้ในแต่ละทิศ จะมีการเน้นลักษณะบ่งต่างในการรู้จำที่ต่างกัน ซึ่งเมื่อนำมาประกอบกันแล้วจะช่วยให้การตัดสินใจ แบบองค์คณะในขั้นสุดท้าย มีความแม่นยำในการรู้จำมากขึ้น

อย่างไรก็ดี ใช่ว่า การใช้ลักษณะบ่งต่างแบบระบุทิศ ที่มีจำนวนมากจะมีผลดีเสมอไป เพราะเราอาจเจอปัญหา คำสาปมิติ ในระดับชั้นที่สูงขึ้นไปอีกหนึ่งมิติ คือจำนวนของชุดของตัวบ่งต่างมีขนาดของมิติที่ใหญ่มากจะทำให้พารามิเตอร์ที่ใช้ในการตัดสินใจของตัวจำแนกมีจำนวนมาก ซึ่งหากต้องการให้ *การจำแนกมีนัยทั่วไป* (generalization) ที่ดี เราจะต้องใช้จำนวนตัวอย่างในการฝึกฝนจำนวนมาก ซึ่งเป็นไปได้ยากในการรู้จำภาพวัตถุ (ภาพใบหน้า หรือ ภาพยานพาหนะทางทหาร) เพื่อหลีกเลี่ยงปัญหา คำสาปมิติ โดยการนำความคิดของการคัดเลือกตัวบ่งต่างมาใช้ ในที่นี้เราจะทำเลื่อนการคัดเลือกขึ้นมาหนึ่งระดับ เป็น การคัดเลือกชุดของตัวบ่งต่าง แทนที่ ตัวบ่งต่าง

ก่อนอื่น ให้เราพิจารณาในรูปที่ 3.1 โดยใช้คำสั่ง MATLAB ในการแก้สมการเชิงวิเคราะห์ เราจะเห็นว่า 2DPCA ที่เน้นแบบจำลองทางแถวนอน (row) จะทำการเข้ารหัสข้อมูลที่หลากหลายของมิติในแนวนอน ในขณะที่ข้อมูลในแนวดิ่งจะถูกสะสมเข้าด้วยกัน จากความรู้เกี่ยวกับความเป็นไปได้ทางชีวภาพของการมองเห็นที่ได้กล่าวมาก่อนหน้านี้แล้ว ข้อมูลในแนวดิ่งซึ่งเป็นข้อมูลที่สำคัญในการรู้จำเมื่อวัตถุมีการหมุนทิศ จะถูกสะสมเข้าด้วยกันทำให้ลักษณะบ่งต่างที่ได้ลดความสำคัญลง ด้วยเหตุนี้ ตัวบ่งต่างที่ได้จากการสกัดด้วย 2DPCA ที่เน้นแบบจำลองทางแถวนอน ควรจะมีประสิทธิ์ภาพในการรู้จำต่ำกว่า ตัวบ่งต่างที่ได้จากการสกัดด้วย 2DPCA ที่เน้นแบบจำลองทางแถวตั้ง

จากตัวอย่างในรูปที่ 3.1 จะพบว่า กำลังงงานของแถวนอนจะถูกสะสมสำหรับ 2DPCA แบบแถวนอน หรือ หากขยายความให้สามารถเข้าใจได้มากขึ้น จะเห็นได้ว่า สมาชิกตัวแรกของเมทริกซ์ความแปรปรวนร่วมเกี่ยวของ 2DPCA แบบแถวนอน จะเท่ากับ กำลังงานของสมาชิกของแถวตั้งของเมทริกซ์ความแปรปรวนร่วมเกี่ยวของ PCA หรือ เท่ากับ การสะสมของค่า อัตตสหสัมพันธ์ (autocorrelation) a1^2+a4^2+a7^2 ส่วนสมาชิกในแนวนอนแรก คอลัมน์ที่สอง จะเป็นผลรวมของสหสัมพันธ์ไขว้ ระหว่างพิกเซลกับพิกเซลที่อยู่ใกล้เคียง

ในลักษณะที่คล้ายคลึงกัน ความแปรปรวนร่วมเกี่ยวของ 2DPCA แบบแถวตั้ง 2DPCA แปลงแนวขวางแบบแถวนอน 2DPCA แปลงแนวขวางแบบแถวตั้ง เพราะผลของการรวมกำลังงานของความแปรปรวนร่วมเกี่ยวของ PCA ใน 2DPCA ในทิศเฉพาะที่ถูกเลือก ที่ทำให้คำสาปมิติ สามารถถูกแก้ไขให้ดีขึ้นได้

```
% MATLAB program to compute PCA, row-based 2DPCA, column-based 2DPCA (column), and its diagonal
>> syms a1 a2 a3 a4 a5 a6 a7 a8 a9 b1 b2 b3 b4 b5 b6 b7 b8 b9
>> A=[a1 a2 a3;a4 a5 a6;a7 a8 a9]
A =
[ a1, a2, a3]
[ a4, a5, a6]
[ a7, a8, a9]

>> C=A(:)*A(:).'                                % conventional PCA
C =
[  a1^2, a1*a4, a1*a7, a1*a2, a1*a5, a1*a8, a1*a3, a1*a6, a1*a9]
[ a1*a4,  a4^2, a4*a7, a4*a2, a4*a5, a4*a8, a4*a3, a4*a6, a4*a9]
[ a1*a7, a4*a7,  a7^2, a7*a2, a7*a5, a7*a8, a7*a3, a7*a6, a7*a9]
[ a1*a2, a4*a2, a7*a2,  a2^2, a2*a5, a2*a8, a2*a3, a2*a6, a2*a9]
[ a1*a5, a4*a5, a7*a5, a2*a5,  a5^2, a5*a8, a5*a3, a5*a6, a5*a9]
[ a1*a8, a4*a8, a7*a8, a2*a8, a5*a8,  a8^2, a8*a3, a8*a6, a8*a9]
[ a1*a3, a4*a3, a7*a3, a2*a3, a5*a3, a8*a3,  a3^2, a3*a6, a3*a9]
[ a1*a6, a4*a6, a7*a6, a2*a6, a5*a6, a8*a6, a3*a6,  a6^2, a6*a9]
[ a1*a9, a4*a9, a7*a9, a2*a9, a5*a9, a8*a9, a3*a9, a6*a9,  a9^2]

>> G1=A.'*A                                     % row-based 2DPCA
G1 =
[     a1^2+a4^2+a7^2,  a1*a2+a4*a5+a7*a8,  a1*a3+a4*a6+a7*a9]
[ a1*a2+a4*a5+a7*a8,   a2^2+a5^2+a8^2,     a2*a3+a5*a6+a8*a9]
[ a1*a3+a4*a6+a7*a9,  a2*a3+a5*a6+a8*a9,      a3^2+a6^2+a9^2]

>> G2=A*A.'                                     % column-based 2DPCA
G2 =
[     a1^2+a2^2+a3^2,  a1*a4+a2*a5+a3*a6,  a1*a7+a2*a8+a3*a9]
[ a1*a4+a2*a5+a3*a6,      a4^2+a5^2+a6^2,  a4*a7+a5*a8+a6*a9]
[ a1*a7+a2*a8+a3*a9,  a4*a7+a5*a8+a6*a9,      a7^2+a8^2+a9^2]

>> DiagA =                                      % construct diagonal matrix A.
[ a1, a2, a3]
[ a5, a6, a4]
[ a9, a7, a8]

>> G3 = DiagA.'*DiagA                           % row-based diagonal PCA
G3 =
[      a1^2+a5^2+a9^2,  a1*a2+a5*a6+a7*a9,  a1*a3+a4*a5+a8*a9]
[ a1*a2+a5*a6+a7*a9,      a2^2+a6^2+a7^2,  a2*a3+a4*a6+a7*a8]
[ a1*a3+a4*a5+a8*a9,  a2*a3+a4*a6+a7*a8,       a3^2+a4^2+a8^2]

>> G4 = DiagA*DiagA.'                           % column-based diagonal PCA
[      a1^2+a2^2+a3^2,  a1*a5+a2*a6+a4*a3,  a1*a9+a7*a2+a8*a3]
[ a1*a5+a2*a6+a4*a3,       a4^2+a5^2+a6^2,  a5*a9+a7*a6+a4*a8]
[ a1*a9+a7*a2+a8*a3,  a5*a9+a7*a6+a4*a8,       a7^2+a8^2+a9^2]
```

รูปที่ 3.1 ความแปรปรวนร่วมเกี่ยว PCA ดั้งเดิม ระนาบราบ ระนาบดิ่ง และ รุ่นถัดมา

คณะผู้วิจัย (Sanguansat, 2007a) พบว่า เราสามารถแปลง เมทริกซ์ความแปรปรวนร่วมเกี่ยวของ PCA ไปเป็น หนึ่งใน เมทริกซ์ความแปรปรวนร่วมเกี่ยวของ 2DPCA แบบต่างๆ ได้ และ ตัวบ่งต่างที่สกัดได้จาก เมทริกซ์ความแปรปรวนที่แปลงนั้นจะมีอำนาจในการตัดสินใจที่ต่างกัน ซึ่งเราจะเรียกชุดของเมทริกซ์ความแปรปรวนร่วมเกี่ยวที่ถูกแปลงนี้ว่า *ความแปรปรวนร่วมเกี่ยวไขว้* (cross image covariance) รูปที่ 3.2 แสดง หนึ่งความเป็นไปได้ในการจัดหมู่ เมทริกซ์ความแปรปรวนร่วมเกี่ยวของ 2DPCA ที่ผันมาจาก เมทริกซ์ความแปรปรวนร่วมเกี่ยวของ PCA จะเห็นได้ว่า หนึ่งใน เมทริกซ์ความแปรปรวนร่วมเกี่ยวของ 2DPCA (corrG1) เกิดจาก การเลื่อนแถวตั้งของเมทริกซ์ความแปรปรวนร่วมเกี่ยวของ PCA แบบเป็นวงกลม (เหมือนกับการแปลงภาพทางขวาง เมื่อเราต้องการสร้าง 2DPCA แบบขวาง แต่ เมทริกซ์ความแปรปรวนร่วมเกี่ยวของ 2DPCA แบบขวาง จะเป็นค่าผลรวมของกำลังงานที่เกิดจากค่าอัตตสัมพันธ์ ในขณะที่การทำการเลื่อนแบบเป็นวงกลม ของความแปรปรวนร่วมเกี่ยวไขว้ นี้จะเน้น ผลรวมของค่ากำลังงานที่เกิดจากสหสัมพันธ์ระหว่างคู่พิกเซลที่ไม่ใช่พิกเซลเดียวกัน)

เราสามารถสร้าง หมู่ของ 2DPCA ขึ้นเป็นเซตขนาดใหญ่ ที่ประกอบ 2DPCA ไขว้ แบบต่างๆ โดยการปรับเลื่อนแถวตั้ง (คอลัมน์) ของเมทริกความแปรปรวนร่วมเกี่ยวของ PCA ไปเรื่อยๆ จากนั้นในแต่ละครั้งทำการหาผมรวมในแนวทแยง $n \times n$ โดยที่ n เป็นความกว้างของภาพวัตถุ ในทำนองเดียวกัน เราอาจทำการปรับเลื่อนแถวนอนของเมทริกความแปรปรวนร่วมเกี่ยวของ PCA ก็จะได้ ชุดของ 2DPCA ไขว้ อีกชุดหนึ่ง ยกตัวอย่างเช่น ในกรณี ของภาพวัตถุมีขนาด 3×3 เราสามารถ 2DPCA ไขว้ แบบต่างๆได้ 8 แบบ ที่แตกต่างจาก 2DPCA แบบแถวนอน หรือ อื่นใดที่แสดงในรูปที่ 3.1 2DPCA ไขว้ของภาพขนาด 3×3 นี้ทั้งหมดแสดงในรูปที่ 3.2 ตามลำดับ

อย่างไรก็ดี 2DPCA ไขว้ที่สร้างขึ้นใหม่นี้ จำเป็นต้องสร้างขึ้นมาจาก เมทริกซ์ความแปรปรวนร่วมเกี่ยวของ PCA ซึ่งถ้าขนาดของภาพวัตถุมีขนาดใหญ่ เมทริกซ์ความแปรปรวนร่วมเกี่ยวของ PCA ก็จะมีขนาดใหญ่มาก ในบางกรณีก็อาจไม่สามารถคำนวณได้ จึงเป็นการยากที่จะสร้าง 2DPCA ไขว้ จาก PCA ในทางปฏิบัติ

โดยความบังเอิญ ผู้วิจัย (Asdornwised, 2008) พบว่า เมทริกซ์ความแปรปรวนร่วมเกี่ยวไขว้สองมิติ เกิดจาก ผลคูณภายใน (inner product) $\mathbf{B}^{\mathbf{T}}\mathbf{A}$ ระหว่างภาพวัตถุต้นแบบ $\mathbf{A}$ กับ เมทริกซ์ภาพที่ผ่านการแปลง $\mathbf{B}$ โดยที่เมทริกซ์ภาพ $\mathbf{B}$ ถูกสร้างจากการแปลงที่เหมาะสมของเมทริกซ์ภาพ $\mathbf{A}$ ดังที่กล่าวมาแล้ว เราเผชิญกับ ความแปรปรวนร่วมเกี่ยวสองมิติ ที่เกิดจากคู่พิกเซลที่ไม่ใช่พิกเซลเดียวกันแทนที่ความแปรปรวนร่วมเกี่ยวสองมิติ โชคดีที่เราสามารถหาการแปลงเมทริกซ์ $\mathbf{B}$ ที่เหมาะสมได้ ซึ่งในที่นี้ สำหรับภาพวัตถุ ขนาด 3×3 หนึ่งในการแปลงที่เหมาะสมสำหรับเมทริกซ์ $\mathbf{B}$ ที่สามารถทำได้ คือโดยทำการเลื่อนแบบเป็นวงกลมต่อแถวนอนของเมทรกิซ์ภาพ $\mathbf{A}$ หนึ่งแถว และทำการเลื่อนแบบเป็นวงกลมเฉพาะแถวนอนที่

เลื่อน ทางขวาของ รูปที่ 3.2 แสดง เมทริกซ์ **B** ที่ผ่านการแปลงแบบหมุนวนเป็นวงกลม แถวนอน และ สมาชิกในแถวนอน

เป็นที่น่าสังเกตว่า ระเบียบวิธีที่นำเสนอในการแปลงเมทริกซ์ B แล้วได้ เมทริกซ์ความแปรปรวนความร่วมเกี่ยวไขว้ของ 2DPCA โดยบังเอิญนั้น เป็นหนึ่งในวิธีที่สามารถทำได้ ผู้วิจัยมีความเชื่อว่าจะมีวิธีที่สามารถแปลงเมทริกซ์ B อย่างอื่น ที่สามารถสร้างความหลากหลายในการสร้าง MCS ที่ดีกว่านี้ได้ ชุดคำสั่งเทียมสำหรับสร้างเมทริกซ์ความแปรปรวมร่วมเกี่ยวไขว้ ได้แสดงไว้ในภาคผนวก ข.

```
>> corrG1 =                                                    >> B1 =
[ a1*a4+a4*a7+a7*a2, a4*a2+a7*a5+a2*a8, a4*a3+a7*a6+a2*a9]     [ a4, a5, a6]
[ a1*a5+a4*a8+a7*a3, a2*a5+a5*a8+a8*a3, a5*a3+a8*a6+a3*a9]     [ a7, a8, a9]
[ a1*a6+a4*a9+a1*a7, a2*a6+a5*a9+a1*a8, a3*a6+a6*a9+a1*a9]     [ a2, a3, a1]

>> corrG2 =                                                    >> B2 =
[ a1*a7+a4*a2+a7*a5, a7*a2+a2*a5+a5*a8, a7*a3+a2*a6+a5*a9]     [ a7, a8, a9]
[ a1*a8+a4*a3+a7*a6, a2*a8+a5*a3+a8*a6, a8*a3+a3*a6+a6*a9]     [ a2, a3, a1]
[ a1*a9+a1*a4+a4*a7, a2*a9+a1*a5+a4*a8, a3*a9+a1*a6+a4*a9]     [ a5, a6, a4]

>> corrG3 =                                                    >> B3 =
[ a1*a2+a4*a5+a7*a8,   a2^2+a5^2+a8^2, a2*a3+a5*a6+a8*a9]      [ a2, a3, a1]
[ a1*a3+a4*a6+a7*a9, a2*a3+a5*a6+a8*a9,   a3^2+a6^2+a9^2]      [ a5, a6, a4]
[   a1^2+a4^2+a7^2, a1*a2+a4*a5+a7*a8, a1*a3+a4*a6+a7*a9]      [ a8, a9, a7]

>>corrG4 =                                                     >> B4 =
[ a1*a5+a4*a8+a7*a3, a2*a5+a5*a8+a8*a3, a5*a3+a8*a6+a3*a9]     [ a5, a6, a4]
[ a1*a6+a4*a9+a1*a7, a2*a6+a5*a9+a1*a8, a3*a6+a6*a9+a1*a9]     [ a8, a9, a7]
[ a1*a4+a4*a7+a7*a2, a4*a2+a7*a5+a2*a8, a4*a3+a7*a6+a2*a9]     [ a3, a1, a2]

>> corrG5 =                                                    >> B5 =
[ a1*a8+a4*a3+a7*a6, a2*a8+a5*a3+a8*a6, a8*a3+a3*a6+a6*a9]     [ a8, a9, a7]
[ a1*a9+a1*a4+a4*a7, a2*a9+a1*a5+a4*a8, a3*a9+a1*a6+a4*a9]     [ a3, a1, a2]
[ a1*a7+a4*a2+a7*a5, a7*a2+a2*a5+a5*a8, a7*a3+a2*a6+a5*a9]     [ a6, a4, a5]

>> corrG6 =                                                    >> B6 =
[ a1*a3+a4*a6+a7*a9, a2*a3+a5*a6+a8*a9,   a3^2+a6^2+a9^2]      [ a3, a1, a2]
[   a1^2+a4^2+a7^2, a1*a2+a4*a5+a7*a8, a1*a3+a4*a6+a7*a9]      [ a6, a4, a5]
[ a1*a2+a4*a5+a7*a8,   a2^2+a5^2+a8^2, a2*a3+a5*a6+a8*a9]      [ a9, a7, a8]

>> corrG7 =                                                    >> B7 =
[ a1*a6+a4*a9+a1*a7, a2*a6+a5*a9+a1*a8, a3*a6+a6*a9+a1*a9]     [ a6, a4, a5]
[ a1*a4+a4*a7+a7*a2, a4*a2+a7*a5+a2*a8, a4*a3+a7*a6+a2*a9]     [ a9, a7, a8]
[ a1*a5+a4*a8+a7*a3, a2*a5+a5*a8+a8*a3, a5*a3+a8*a6+a3*a9]     [ a1, a2, a3]

>> corrG8 =                                                    >> B8 =
[ a1*a9+a1*a4+a4*a7, a2*a9+a1*a5+a4*a8, a3*a9+a1*a6+a4*a9]     [ a9, a7, a8]
[ a1*a7+a4*a2+a7*a5, a7*a2+a2*a5+a5*a8, a7*a3+a2*a6+a5*a9]     [ a1, a2, a3]
[ a1*a8+a4*a3+a7*a6, a2*a8+a5*a3+a8*a6, a8*a3+a3*a6+a6*a9]     [ a4, a5, a6]
```

รูปที่ 3.2 ความแปรปรวนร่วมเกี่ยวไขว้ สองมิติ และเมทริกซ์ B ที่ใช้ในการแปลง

3.2 การสร้างคืนปริภูมิย่อยสองมิติความละเอียดสูงยวดยิ่ง

จากความรู้คณิตศาสตร์พื้นฐานที่ได้กล่าวถึงในบทก่อนหน้านี้ เราจะสามารถเขียนการแทนรูปสัญญาณ โดยการพิจารณาการฉายด้วย PCA โดยสมการเชิงเส้น

$$\tilde{\mathbf{x}} = \boldsymbol{\Theta}\mathbf{v} + \tilde{\mathbf{e}}_{\mathbf{x}} \tag{21}$$

โดยที่ $\tilde{\mathbf{x}}$ แทนรูปเมทริกซ์ของภาพใบหน้าทั้งหมดที่อยู่ในรูปการต่อกันของเมทริกซ์ภาพใบหน้าดั้งเดิม $\{\theta_1,\cdots,\theta_d\}$ เป็นชุดของ เวกเตอร์เฉพาะ (eigenvectors) ที่มีค่ามากที่สุดเรียงลำดับจากมากไปหาน้อย ที่ต่อกันเป็นเมทริกซ์เฉพาะ $\boldsymbol{\Theta}$ และ $\mathbf{v}$ เป็น เมทริกซ์ของสัมประสิทธิ์ที่มีจำนวนที่สอดคล้องกับ จำนวนเวกเตอร์เฉพาะ หรือ ตัวบ่งต่าง ที่ได้จากการฉายภาพ HR ลงบน $\boldsymbol{\Theta}$ ซึ่งในที่นี้จะเรียกว่า *เมทริกซ์ส่วนประกอบมุขสำคัญ* (principal component matrix) โดยเงื่อนไขที่ใช้เพื่อหาคำตอบในสมการ (21) ได้กล่าวถึงมาแล้วในบทที่ 2

3.2.1 ทางเลือกแบบจำลองภาพ

ในบทย่อยนี้ เราจะนำเสนอทางเลือกในการจำลองภาพอย่างง่ายที่เหมาะสมต่อการสร้างคืนปริภูมิย่อยความละเอียดสูงยวดยิ่ง โดยเราจะกำหนดให้ ภาพ LR และ HR มีความสัมพันธ์กัน (Vijay, 2008) ดังนี้

$$\tilde{\mathbf{f}}_k = L_k \tilde{\mathbf{x}} R_k + \tilde{\mathbf{n}}_k, \quad 1 \le k \le p \tag{22}$$

โดยที่ $\boldsymbol{p}$ เป็นจำนวนของเฟรมที่พิจารณา L_k, R_k เป็นเมทริกซ์ของการสุ่มสัญญาณต่ำ และ $\tilde{\mathbf{f}}_k$ และ $\tilde{\mathbf{x}}$ เป็นเมทริกซ์ภาพที่ได้จากเฟรมภาพ LR และ HR ที่ k^{th} ตามลำดับ เป็นที่น่าสังเกตว่า ตัวพร่าเกาส์สองมิติ นั้นสามารถแทนที่ได้ด้วย เมทริกซ์ L_k และ R_k ที่แยกออกจากกันได้สองตัว (ดูรายละเอียดเพิ่มเติม ในบทย่อย ที่ 3.1.1 ข้างล่างนี้) เป็นที่น่าสังเกตว่า ภาพใบหน้า LR และ HR ที่ใช้ในบทนี้ จะอยู่ในรูปเมทริกซ์ภาพดั้งเดิม ไม่มีการแปลงเมทริกซ์เป็นเวกเตอร์แบบ *เรียงคำ* (lexicography) เหมือนกับ ระเบียบวิธีที่ใช้ใน PCA ตามสมการที่ (1)

3.2.2 การทำให้ราบเรียบด้วยตัวทำพร่าเกาส์

ตัวดำเนินการทำให้ราบเรียบเกาส์ (Gsmooth, n.d.) เป็นตัวดำเนินการ *สังวัตนาการ* (convolution) สองมิติ ที่นิยมใช้ในการทำพร่าภาพ หรือ การขจัดรายละเอียด หรือ สัญญาณรบกวน ลักษณะการทำงานมีความคล้ายคลึงกับ ฟิวเตอร์เฉลี่ย (mean filter) เพียงแต่ใช้แก่นในการดำเนินการที่แตกต่างกันไป ซึ่งก็คือ รูปสัญญาณหลังค่อมที่อยู่ในรูปแบบของเกาส์ (ระฆัง)

$$\frac{1}{273}\begin{bmatrix} 1 & 4 & 7 & 4 & 1 \\ 4 & 16 & 26 & 16 & 4 \\ 7 & 26 & 41 & 26 & 7 \\ 4 & 16 & 26 & 16 & 4 \\ 1 & 4 & 7 & 4 & 1 \end{bmatrix}$$

รูปที่ 3.3 การประมาณเต็มหน่วยสำหรับฟังก์ชันเกาส์

อันที่จริงการทำให้ราบเรียบแบบเกาส์ ซึ่งเป็นเหมือน *ฟังก์ชันกระจายจุด* (point spreading function) โดยเป็นการสังวัตนาการระหว่างเมทริกซ์ภาพ ที่เก็บไว้เป็นพิกเซลเต็มหน่วย กับ เมทริกซ์ของฟังก์ชันเกาส์ประมาณที่เก็บไว้แบบเต็มหน่วย เช่นกัน ในทางทฤษฎีฟังก์ชันเกาส์จะมีค่าไม่เป็นศูนย์ทุกที่ ซึ่งทำให้การทำสังวัตนาการต้องทำเป็นอนันต์ ซึ่งเป็นไปไม่ได้ในทางปฏิบัติ ในทางปฏิบัติเราอาจประมาณให้ค่าของฟังก์ชันเกาส์เท่ากับศูนย์ หากของตำแหน่งที่เกินกว่าสามเท่าของความแปรปรวนจาก *ค่าเฉลี่ยกลาง* (mean) แก่นของการทำให้ราบเรียบจะถูกทำให้สั้นลง รูปที่ 3.3 แสดงแก่นสังวัตนาการจำนวนเต็มที่ใช้ประมาณฟังก์ชันเกาส์ที่มี ความเบี่ยงเบน (σ) เท่ากับ 1.0

แก่นสังวัตนาการข้างต้นนี้ สามารถนำใช้ในการสังวัตนาการสองมิติกับเมทริกซ์ภาพในการสร้างคืนภาพความละเอียดสูงสุดนี้ เราสามารถนำมาสร้างเป็นเมทริกซ์ แล้วนำมาคูณกับภาพ HR ที่ถูกแปลงเป็นเวกเตอร์แบบเรียงคำ ซึ่งจะได้ผลลัพธ์เช่นเดียวกับการสังวัตนาการเช่นกัน

อย่างไรก็ดี การแปลงเมทริกซ์ทำพร่าจากแก่นทำพร่าเกาส์สองมิติ นั้นเหมาะสำหรับการสร้างคืนภาพความละเอียดสูงยวดยิ่ง และ สร้างคืนส่วนประกอบมุขสำคัญความละเอียดสูงยวดยิ่งหนึ่งมิติ เท่านั้น แต่เป็นแบบจำลองทางคณิตศาสตร์ที่ไม่เหมาะกับการสร้างคืนส่วนประกอบมุขสำคัญความละเอียดสูงยวดยิ่ง สองมิติ

เนื่องด้วย การสังวัตนาการภาพ โดยการใช้แก่นทำพร่าเกาส์ที่ถูกประมาณอย่างเต็มหน่วยข้างต้น สามารถกระทำได้รวดเร็วขึ้น โดยอาศัยการคำนวณด้วย ผลเทรเซอร์ (tensor product) ของแนวแกน X และ Y ที่แยกออกจากันได้ ทั้งนี้เนื่องมาจากว่า การสังวัตนาการภาพสองมิติด้วยฟังก์ชันเกาส์สองมิติ นั้นเป็นการดำเนินการแบบหนึ่งเดียวในทุกทิศทาง (2D isotropic)

.006	.061	.242	.383	.242	.061	.006

รูปที่ 3.4 แก่นสังวัตนาการหนึ่งมิติ เพิ่มความเร็วในการคำนวณการสังวัตนาการสองมิติ

ด้วยเหตุนี้ การสังวัตนาการสองมิติ สามารถดำเนินการได้ด้วย การสังวัตนาการหนึ่งมิติกับภาพทางทิศ X จนหมด แล้วจึงทำการสังวัตนาการหนึ่งมิติกับเอาท์พุตที่ได้มาก่อนนี้ ทางทิศ Y หรือ อีกนัยหนึ่ง เนื่องจาก ฟังก์ชันเกาส์ ที่เป็นดำเนินการที่สมมาตรแบบกลมโดยสมบูรณ์ จึงสามารถแยกย่อยออกได้ดังกล่าว รูปที่ 3.4 แสดงแก่นหนึ่งมิติ ที่สามารถนำมาใช้เพื่อให้ผลการสังวัตนาการได้ผลเช่นเดียวกับการใช้แก่นทำพร่าเกาส์สองมิติ ที่แสดงในรูปที่ 5.3 เป็นที่น่าสังเกตว่า เวกเตอร์มีขนาด 7 เมื่อเราทำการเปลี่ยนขนาดด้วย 273 ให้เป็นจำนวนเต็มที่ใกล้ที่สุด และ กำจัดพิกเซลที่ใกล้ศูนย์ที่อยู่ที่ขอบภาพเนื่องจากมีค่าใกล้กับศูนย์ เราจะได้เมทริกซ์ขนาด 5×5 แทนที่เมทริกซ์ขนาด 7×7 สำหรับ เวกเตอร์ในการทำพร่าทาง Y จะมีค่าเท่ากัน เพียงแต่จะวางในแนวตั้ง

มีข้อน่าสังเกต สองประการ ประการแรก เราสามารถทำพร่าด้วยตัวทำเกาส์ที่ความแปรปรวนสูงได้โดย เปลี่ยนเป็นการทำสังวัตนาการภาพด้วยตัวทำพร่าเกาส์ที่มีความแปรปรวนน้อยจำนวนหลายๆครั้ง ซึ่งถึงแม้ว่าจะเป็นการคำนวณที่ซับซ้อนแต่เราสามารถทำได้โดยการคำนวณแบบขนาน ประการที่สอง ฟังก์ชันเกาส์ เป็นที่สนใจของนักชีววิทยาที่คำนวณได้ (computational biologist) เนื่องมาจากความเป็นไปได้ทางชีวภาพ เช่น พบเซลบางชนิดที่อยู่ในทางผ่าน หรือ วิถี ของการมองเห็นที่มีในสมอง ที่สามารถน่าจะมีผลตอบสนองที่เหมือนกับฟังก์ชันประมาณของเกาส์

โดยการแยกออกจากกันของแก่นการทำพร่า เราจึงสามารถสร้างแบบจำลองทางคณิตศาสตร์ เป็นทางเลือกใหม่ โดย L_k และ R_k เป็นตัวดำเนินการทำพร่า และ การสุ่มสัญญาณลดลง ทางซ้าย และ ขวา ตามลำดับ ดังแสดงในสมการที่ (22)

3.2.3 ระเบียบวิธีการสร้างคืนที่นำเสนอ

ดังนั้น เราจะสามารถสร้างสมการความสัมพันธ์ความละเอียดสูงยวดยิ่งของปริภูมิ ได้ใหม่ดังนี้

$$\mathbf{\Gamma}\hat{\mathbf{v}}_k + \mathbf{e}_{\tilde{f}_k} = L_k \mathbf{\Theta v} R_k + L_k \mathbf{e}_x R_k + \tilde{\mathbf{n}}_k \qquad (23)$$

โดยที่ $\{\Gamma_1,\cdots,\Gamma_d\}$ เป็น เวกเตอร์เฉพาะที่มีค่ามากที่สุดเรียงลำดับจากมากไปหาน้อย จำนวน *d* ตัว ที่ประกอบขึ้นรวมกันเป็นเมทริกซ์ $\mathbf{\Gamma}$ และ $\hat{\mathbf{v}}_k$ เมทริกซ์ของสัมประสิทธิ์ที่มีจำนวนที่สอดคล้องกับ จำนวนเวกเตอร์เฉพาะ หรือ ตัวบ่งต่าง ที่ได้จากการฉายภาพ LR ลงบน $\mathbf{\Gamma}$

โดยไม่มีผลต่อนัยสำคัญ เราได้ว่า

$$\mathbf{\Gamma}^T \mathbf{e}_{f_k} = 0 \tag{24}$$

และ

$$\mathbf{\Gamma}^T \mathbf{\Gamma} = 1 \tag{25}$$

ไม่เป็นการยาก ที่จะผันสมการต่อไปนี้

$$\hat{\mathbf{v}}_k = \mathbf{\Gamma}^T L_k \mathbf{\Theta} \mathbf{v} R_k + \mathbf{\Gamma}^T L_k \mathbf{e}_x R_k + \mathbf{\Gamma}^T \tilde{\mathbf{n}}_k \tag{26}$$

เป็นที่น่าสังเกตว่า $\hat{\mathbf{v}}_k$ ในที่นี้ เป็นเมทริกซ์บ่งต่าง ซึ่งไม่เหมือนกับ $\hat{\mathbf{a}}_k$ ที่มีลักษณะเป็นเวกเตอร์บ่งต่าง ด้วยเหตุนี้จึงมีซับซ้อนเล็กน้อยต่อการแก้ปัญหาผกผัน ของตัวบ่งต่าง ที่เป็น เมทริกซ์ปริภูมิย่อยความละเอียดสูงยวดยิ่ง $\mathbf{v}_k$ โดยการใช้ตัวดำเนินการทางเวกเตอร์ที่นำเสนอโดย Kumar and Schott (Kumar, 2008 & Schott, 2005),

สมการที่ (26) จะสามารถเขียนได้ใหม่เป็น

$$\hat{\beta}_k = \Xi_k \beta + \eta_k \tag{27}$$

โดยที่ $\hat{\beta}_k = vec(\hat{\mathbf{v}}_k)$ $\beta_k = vec(\mathbf{v})$ $\Xi_k = R_k^T \otimes \mathbf{\Gamma}^T L_k \mathbf{\Theta}$ และ $\eta_k = vec(\mathbf{\Gamma}^T L_k \mathbf{e}_x R_k + \mathbf{\Gamma}^T \tilde{\mathbf{n}}_k)$ ในที่นี้ $\otimes$ เป็นตัวดำเนินการ Kronecker ดังนี้ เราจะสามารถแก้ปัญหาเมทริกซ์ตัวบ่งต่างสองมิติ ที่อยู่ในระดับปริภูมิย่อยเฉพาะ โดยระเบียบวิธีที่คล้ายคลึงกับ สมการที่ (15) และ (18) ได้ดังนี้

$$\boldsymbol{\beta} = \boldsymbol{\Xi}^{\mathbf{T}} (\boldsymbol{\Xi}\boldsymbol{\Xi}^{\mathbf{T}} + \gamma \mathbf{I})^{-1} \hat{\boldsymbol{\beta}} \tag{28}$$

โดยที่ γ เป็นนิพจน์ในการเรกูราไรซ์ หลังจากเราแปลง $\boldsymbol{\beta}$ กลับไปเป็นเมทริกซ์ เราจะได้เมทริกซ์ตัวบ่งต่างความละเอียดสูงยวดยิ่งตามต้องการ

3.3 MCS สำหรับการรู้จำวัตถุ

บทนี้ เราจะกล่าวถึง ระเบียบที่เราพัฒนาขึ้น เพื่อใช้กับ ชุดของ 2DPCA ไขว้ ที่สร้างขึ้นด้วยระเบียบวิธีในบทที่ 3.2 ได้แก่

1. ระเบียบวิธี A ซึ่งใช้แนวคิดที่จะสร้างตัวจำแนกที่เป็นอิสระต่อกัน จำนวนหนึ่งเพื่อจะได้รวมการตัดสินใจในขั้นสุดท้าย เพื่อให้ผลในการรู้จำดีขึ้น
2. ระเบียบวิธี B ซึ่งจะทำการคัดเลือก 2DPCA เฉพาะที่เมื่อรวมการตัดสินใจเข้าด้วยกันในขั้นสุดท้ายแล้วจะให้ผลในการรู้จำดีขึ้น ในการคัดเลือก 2DPCA เฉพาะที่คาดว่าจะดีต่อการรู้จำนั้น เราใช้แนวคิดแบบฟิวเตอร์

รายละเอียดของระเบียบวิธีทั้งสอง จะได้แสดง ดังต่อไปนี้

3.3.1 ระเบียบวิธี A: ปริภูมิย่อยสุ่ม

ระเบียบวิธีปริภูมิย่อยสุ่ม (Random Subspace Method: RSM) (Ho, 1998) เป็นหนึ่งในระเบียบวิธีของ MCS เช่นเดียวกับ Bagging และ Boosting แต่มีข้อแตกต่างอยู่ที่ทั้ง Bagging และ Boosting ไม่มีการลดขนาดของมิติ เพื่อช่วยในการตัดสินใจ ในขณะที่ RSM มีการลดขนาดมิติร่วมอยู่ด้วย การที่ไม่มีการลดขนาดของมิติจะทำให้ตัวจำแนกแต่ละตัวที่อยู่ในคณะกรรมการตัดสินใจมีความแปรปรวนในการตัดสินใจสูง และมีความเอนเอียงต่ำ ซึ่งอาจไม่เหมาะกับการรู้จำภาพวัตถุที่เราพยายามหลีกเลี่ยง คำสาปมิติ อยู่ (Bagging จะทำการเลือกตัวอย่างที่ใช้ในการฝึกฝนอย่างสุ่ม เพื่อให้ตัวจำแนกทำการเรียนรู้ ในกรณีเนื่องจาก จำนวนตัวอย่างมีน้อยอยู่แล้ว การเลือกให้น้อยลงไปอีกจะทำให้เกิด *การจำ* (memory) ตัวอย่างที่ใช้ฝึกฝน และ เนื่องจากมิติมีขนาดใหญ่ ทำให้ตัวจำแนกต้องมีความซับซ้อนสูงตามไปด้วย ในขณะที่ Boosting จะทำการปรับการถ่วงน้ำหนักการกระจายตัวของตัวอย่างที่ได้รับการฝึกฝนแล้ว) ด้วยเหตุนี้ RSM จึงสามารถใช้ประโยชน์จากการลดขนาดมิติได้ดี เนื่องจาก RSM สามารถสร้างคณะของตัวจำแนก ซึ่งแต่ละตัวมีความเป็นอิสระต่อกัน เราจึงสามารถรวมการตัดสินใจได้อย่างมีระเบียบแบบแผนง่ายๆ เช่น การลงคะแนนเสียง หรือ กฎการบวก เป็นต้น

โดยสรุป ประโยชน์ของการใช้ RSM มีดังต่อไปนี้

1. สามารถหลีกเลี่ยง คำสาปมิตี ด้วยการลดขนาดมิติ
2. เหมาะกับจำนวน ตัวอย่างในการฝึกฝน น้อยๆ เช่น การรู้จำภาพวัตถุ
3. ตัวจำแนกแบบสมาชิกใกล้เคียงที่สุด (nearest neighbor classifier) ที่นิยมใช้ในการรู้จำวัตถุนั้น ค่อนข้างอ่อนไหวต่อ การกระจายตัวอย่างหลวมมาก (sparsity) ของปัญหาที่มีขนาดมิติใหญ่มาก
4. ไม่การหาคำตอบแบบวนซ้ำ หรือ การปีนเขา (hill climbing) จึงไม่ต้องกลัวว่าจะมีการตกหลุมเล็งเลิศท้องถิ่น (local optimum)
5. RSM มีสมรรถนะที่ดีกว่า การรู้จำด้วยตัวจำแนกแข็งแรง (strong classifier) เพียงตัวเดียว

ด้วยเหตุนี้ RSM จึงสามารถหลีกเลี่ยงปัญหา คำสาปมิติ แลt SSS รวมทั้งความแม่นยำในการรู้จำที่เหมาะสมกับปัญหาที่มีจำนวนตัวอย่างในการฝึกฝนน้อย และ มีขนาดมิติใหญ่มาก RSM ได้ถูกนำมาใช้กับ 2DPCA สำหรับการรู้จำใบหน้า (Chawla, 2005; Nguyen, N. 2007; Wang, 2006; Sanguansat, 2007b & Sanguansat, 2007c) จากงานวิจัยข้างต้น พบว่า RSM ไม่เหมาะกับการใช้ร่วมกับ PCA เพราะเมื่อเราทำการเลือก ค่าบ่งต่างที่ได้จากการฉายภาพลงบนเวกเตอร์เฉพาะแบบสุ่มเพื่อสร้างเป็นตัวบ่งต่างจำนวนมากนั้น แต่ข่าวสารที่มีอยู่

ในแต่ละสมาชิก (ค่าบ่งต่าง) ของ PCA นั้นไม่เท่าเทียมกัน เวกเตอร์เฉพาะที่มี ค่าเฉพาะ (eigenvalue) มากกว่า โดยส่วนมากจะข่าวสารที่สำคัญต่อการรู้จำมากกว่า หากตัวบ่งต่าง ที่ตรงกับเวกเตอร์เฉพาะที่มีค่ามากนั้นไม่ถูกเลือก ตัวจำแนกที่ใช้ค่าบ่งต่างนั้นในการตัดสินใจ จะไม่มีข่าวสารที่ดีพอในการตัดสินใจให้แม่นยำ

ต่างกับ การประยุกต์ใช้กับ 2DPCA ที่เราจะได้ ค่าบ่งต่างเป็นเมทริกซ์ ซึ่งหากเราเลือก คอลัมน์ของตัวบ่งต่าง 2DPCA แบบสุ่ม คอลัมน์ของตัวบ่งต่างไม่ขึ้นกับเวกเตอร์เฉพาะ ด้วย เหตุนี้ RSM จึงเหมาะที่จะใช้งานร่วมกับ 2DPCA การใช้งานของ RSM เกิดขึ้นจากการ กระทำเพื่อทำการเลือกค่าของปริภูมิย่อยของตัวบ่งต่างอย่างสุ่มของข้อมลที่ใช้ในการฝึกฝน โดย ปกติเราจะทำการสุ่มแบบเอกรูป เพื่อสร้างเซตย่อย โดยที่ ตัวบ่งต่างใหม่ที่ได้จะมีขนาด มิติน้อยกว่าขนาดมิติของงตัวบ่งต่างดั้งเดิม เนื่องจากข่าวสารที่ปรากฏต่อตัวจำแนกมี ขนาดมิติที่ลดลง ตัวจำแนกที่ได้จากการเรียนรู้จึงเป็น *ตัวจำแนกอย่างอ่อน* (weak classifier) อย่างไรก็ดี ตัวจำแนกอย่างอ่อน ตัวเดียวอาจให้สมรรถนะที่ด้อย แต่ผลการรู้จำจะดีขึ้นหากมี การรวมการตัดสินใจของเซตย่อยของตัวจำแนกอย่างอ่อนนี้ ในขั้นสุดท้ายของการทำนายชุด ของตัวอย่างทดสอบ

RSM จึงได้ประโยชน์จากการใช้ปริภูมิย่อย ทั้งในการสร้างและการรวบรวมตัวจำแนก โดยเฉพาะเมื่อจำนวนของตัวอย่างที่นำมาฝึกฝนมีจำนวนน้อยเมื่อเทียบกับขนาดมิติ โดย การการสร้างเซตย่อยด้วยปริภูมิย่อยแบบสุ่มนี้ เราจะสามารถแก้ปัญหา SSS ได้ ขนาดมิติ ของปริภูมิย่อยที่เล็กกว่าปริภูมิเดิม ในขณะที่จำนวนตัวอย่างที่ใช้ในการฝึกฝนเท่าเดิม ด้วยเหตุนี้ จำนวนตัวอย่างสัมพัทธ์ที่ใช้ในการเรียนรู้จึงดูเหมือนมีมากขึ้น แม้ทว่าตัวจำแนกแต่ ละตัวจะร่วมกันใช้ข้อมูลบางส่วนที่ซ้ำซ้อนกัน แต่การรวมการตัดสินใจในปริภูมิย่อยที่เป็นอย่าง สุ่มจะให้ผลการตัดสินใจขั้นสุดท้ายที่ดีขึ้นกว่า ตัวจำแนกอย่างเข้มแข็ง ที่เรียนรู้ด้วยข้อมูลที่มี ขนาดมิติคงเดิม

ชุดคำสั่งเทียมสำหรับสร้างระบบรู้จำแบบหลายตัวด้วยวิธีปริภูมิย่อยสุ่มสำหรับ 2DPCA แนวขวาง ได้แสดงไว้ในภาคผนวก ข.

3.3.2 ระเบียบวิธี B: ชุดปริภูมิย่อยเฉพาะ

แทนที่เราจะเลือก มิติของตัวบ่งต่างอย่างสุ่ม หรือ เลือกสมาชิกในเซตของ 2DPCA ไขว้ อย่างสุ่ม เราอาจเลือกทางเลือกใหม่ โดยการใช้ สมการที่ (17) (19) หรือ (20) เพื่อใช้ใน เพื่อคำนวณ ราคา (cost) สำหรับ การคัดเลือกตัวบ่งต่างแบบฟิวเตอร์ (filter method)

ชุดคำสั่งเทียมสำหรับสร้าง ระบบรู้จำแบบหลายตัวด้วยระเบียบวิธีคัดเลือกตัวบ่งต่าง 2DPCA ไขว้ ได้แสดงไว้ในภาคผนวก ข.

4. การวิเคราะห์ส่วนประกอบมุขสำคัญด้วยผลการแปลงเวฟเลตระบุทิศทาง

เพื่อสมรรถนะการเลือกทิศทาง ปริภูมิใหม่ที่สนใจควรเป็นปริภูมิที่มีข่าวสารที่สำคัญต่อการรู้จำ ลักษณะบ่งต่างที่ได้จากการสกัดควรจะอ่อนไหวต่อทิศทางที่ถูกระบุ และไม่อ่อนไหวในทิศทางที่ไม่ถูกระบุ ยกตัวอย่างเช่น ปริภูมิระบุทิศ ที่ควรจะถูกเลือกควรจะต้องคงทนต่อความสัมพันธ์ของส่วนของภาพใบหน้าที่เกียวข้องกับผิวหน้า ทั้งในแนวระนาบและแนวดิ่งในบริเวณเล็กๆ ปริภูมิที่ถูกเลือกอาจสร้างขึ้นโดย 2DPCA ตามระเบียบที่กล่าวมาก่อนหน้านี้

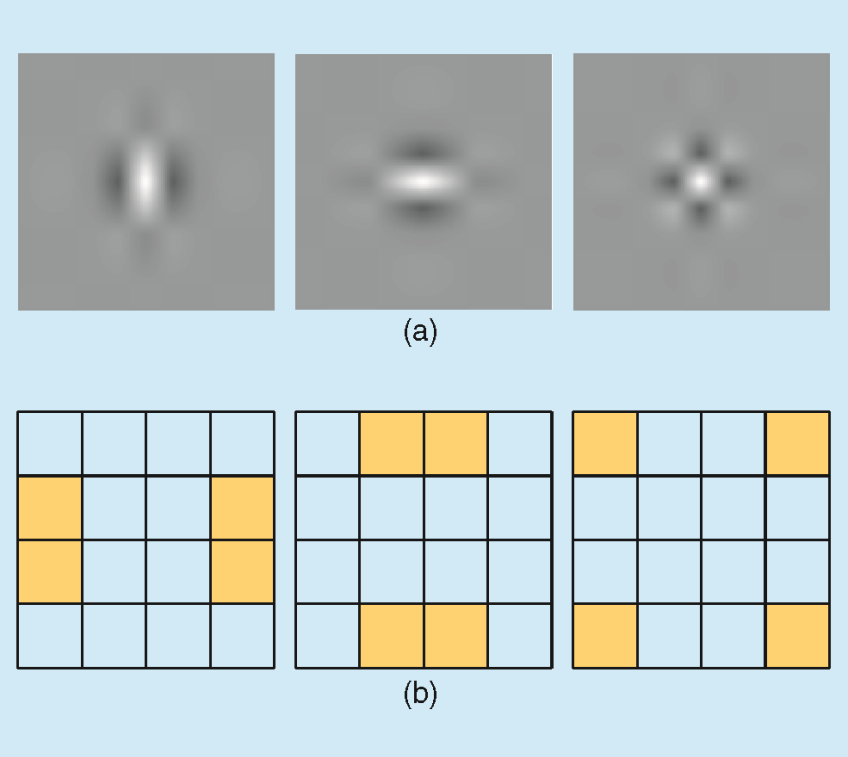
อย่างไรก็ดี มีอีกทางเลือกหนึ่งในการสกัดลักษณะบ่งต่างที่มีข่าวสารสำคัญที่ขึ้นกับทิศทาง โดยดำเนินการผ่านผลการแปลงเวฟเลตที่สามารถระบุทิศก่อน จากนั้นเราจึงทำการสกัดผลบ่งต่างด้วย PCA หรือ 2DPCA ที่ไม่ต้องระบุทิศทาง ซึ่งในที่สุดเราจะได้ลักษณะบ่งต่างที่มีทิศทาง เมื่อได้ ตัวบ่งต่าง จากแถบความถี่ย่อยที่มีทิศทางแล้ว เราอาจทำการคัดเลือกตัวบ่งต่างด้วย สมการที่ (17) (19) หรือ (20) เพื่อคำนวณ ราคา (cost) สำหรับใช้ร่วมกับการคัดเลือกตัวบ่งต่างแบบฟิวเตอร์ (filter method)

4.1 Dual-Tree Wavelet ชนิดจำนวนจริง

แม้ทว่า ฐานหลักของเวฟเลตจะดีที่สุดโดยนัย สำหรับสัญญาณขนาดหนึ่งมิติโดยส่วนใหญ่ แต่สำหรับ ผลการแปลงเวฟเลตสองมิติ อาจจะไม่สามารถคงซึ่งคุณสมบัติที่ดีที่สุดสำหรับภาพธรรมชาติทั่วไปได้ เหตุผลน่าจะอยู่ที่ว่า ผลการแปลงเวฟเลตสองมิติแบบแยกออกจากกันได้สามารถแทนรูปภาวะเอกฐานของจุดได้อย่างมีประสิทธิภาพ แต่จะไม่มีประสิทธิภาพต่อภาวะเอกฐานในแนวเส้นตรง หรือ เส้นโค้ง (เส้นขอบ) ด้ยเหตุนี้จึงได้มีงานวิจัยหลากหลายที่วิจัย ผลการแปลงแบบหลายระดับความละเอียดสองมิติ ที่สามารถแทนรูปขอบภาพได้อย่างมีประสิทธิภาพกว่า ผลการแปลงเวฟเลตสองมิติแบบแยกออกจากกันได้ หนึ่งในเวฟเลตที่ระบุทิศทางได้ คือ dual-tree wavelet ชนิดจำนวนจริง (real dual-tree wavelet) ซึ่งเราจะได้ขยายความ และ แสดงชุดคำสั่ง MATLAB ที่จะใช้ในงานวิจัยนี้ ดังนี้

ก่อนอื่นขอให้เราพิจารณา ผลการแปลงเวฟเลตแบบแยกออกจากกันได้ เราจะพบว่ามีเวฟเลตที่มีทิศ 3 เวฟเลต คือ

$$\begin{aligned} \psi_1(x,y) &= \phi(x)\,\psi(y) \qquad \text{(LH wavelet)}, \\ \psi_2(x,y) &= \psi(x)\,\phi(y) \qquad \text{(HL wavelet)}, \\ \psi_3(x,y) &= \psi(x)\,\psi(y) \qquad \text{(HH wavelet)}. \end{aligned} \tag{29}$$



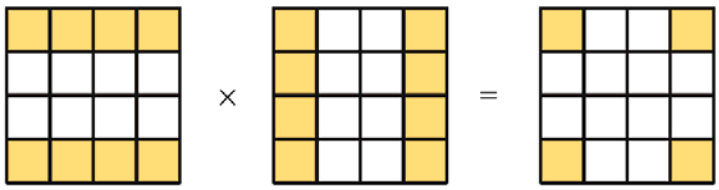
รูปที่ 4.1 เวฟเลตที่ใช้ใน ผลการแปลงเวฟเลตเต็มหน่วย รูป (a) แสดงเวฟเลตที่อยู่ในปริภูมิตำแหน่งของแถบความถี่ย่อย LH HL และ HH รูป (b) แสดงผลตอบสนองทางความถี่ มีจุดศูนย์กลางอยู่ที่กลางภาพ และ HH มีสิ่งแปลกปน

โดยที่ LH เป็นผลคูณของ ฟังก์ชันกรองผ่านต่ำ $\phi(.)$ ไปทางมิติที่หนึ่ง และ $\psi(.)$ ฟังก์ชันกรองผ่านสูง ไปทางมิติที่สอง $\phi(.)$ เรียกอีกอย่างหนึ่งว่า ฟักชันสเกลลิงหนึ่งมิติ ส่วน $\psi(.)$ เรียกอีกอย่างหนึ่งว่า ฟังก์ชันเวฟเลตหนึ่งมิติ

ในทำนองเดียวกัน แถบความถี่ย่อย HL และ HH ก็เป็นผลคูณของฟังก์ชันสเกลลิง และเวฟเลต ดังสมการที่ (29) ตามลำดับ รูปที่ 4.1 แสดง ฟังก์ชันเวฟเลตในปริภูมิตำแหน่ง และผลตอบสนองทางความถี่ ของมัน เป็นๆที่น่าสังเกตว่า แถบความถี่ HH ให้ผลตอบสนองที่มี สิ่งแปลกปนแบบตารางหมากรุก เกิดขึ้น สิ่งแปลกปนนี้เป็นสิ่งที่แสดงให้เห็นถึงความไม่มีประสิทธิภาพของผลการแปลงเวฟเลตแบบแยกจากกันได้ หรือ อีกนัยหนึ่ง ผลการแปลงเวฟเลตแบบแยกจากกันได้ ไม่สามารถแทนรูปในทิศ 45 องศา ได้ดีนั่นเอง

เหตุผลที่ผลการแปลงเวฟเลตแบบแยกจากกันได้ ไม่สามารถแทนรูปภาพในแนวแกน 45 องศา ได้ดี น่าจะเกิดจากว่า $\psi(.)$ เป็นฟังก์ชันจริง ดังนั้นสเปกตรัมของฟังก์ชันจึงเป็นแบบ สองข้าง (two-sided) จึงไม่สามารถหลีกเลี่ยงได้ที่จะเกิด สัญญาณความถี่กลาง ขึ้นในทั้งสี่มุม ของระนาบสองมิติ เนื่องจากการรบกวนทางความถี่ที่เกิดขึ้น จึงทำให้เห็นความไม่คมชัดในทางปริภูมิตำแหน่ง ผลเทนเซอร์ ของ HH แสดงในไดอะแกรม ต่อไป นี้

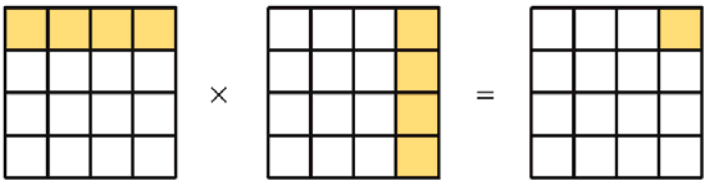
และ ไดอะแกรมของสเปกตรัมของผลคูณของ HH ที่มีทิศ แสดงข้างล่างนี้



สำหรับ dual-tree wavelet ชนิดจำนวนจริงนั้น Selesnick และคณะได้เสนอ เวฟเลตที่เป็น ฟังก์ชันวิเคราะห์ (analytic function) หรือเป็น ฟังก์ชันเชิงซ้อน ที่มีสเปกตรัมทางความถี่เป็นแบบข้างเดียว (one-sided spectrum) ผลคูณเทรเซอร์ แบบเดียวกับ ผลการแปลงเวฟเลตแบบแยกจากกันได้ แสดงดังสมการที่ (30)

$$\begin{aligned}\psi(x,y) &= [\psi_h(x) + \mathrm{j}\,\psi_g(x)]\,[\psi_h(y) + \mathrm{j}\,\psi_g(y)] \\ &= \psi_h(x)\,\psi_h(y) - \psi_g(x)\,\psi_g(y) \\ &\quad + \mathrm{j}\,[\psi_g(x)\,\psi_h(y) + \psi_h(x)\,\psi_g(y)].\end{aligned} \tag{30}$$

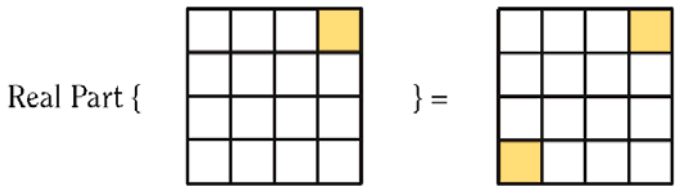
และ ไดอะแกรมของสเปกตรัมของผลคูณ ที่มีสเปกตรัมแบบข้างเดียว แสดงข้างล่างนี้



เพื่อให้เกิดทิศของเวฟเลตโดยสมบรูณ์ Selesnick และคณะ เสนอให้ทำการหาค่าจำนวนจริงของสมการที่ (30) ทำให้เกิดสเปกตรัมแบบสองข้างขึ้น ซึ่งนับว่าเป็นความคิดที่ฉลาดมาก

$$\text{Real Part}\{\psi(x,y)\} = \psi_h(x)\,\psi_h(y) - \psi_g(x)\,\psi_g(y). \tag{31}$$

และ ไดอะแกรมของสเปกตรัมของผลคูณของ HH ที่หันทิศ 45 องศา แสดงข้างล่างนี้



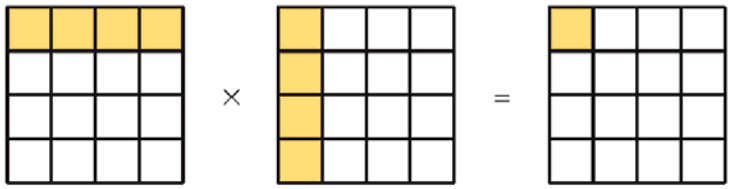
ในทำนองเดียวกัน เราจะได้ เวฟเลตในทิศที่หันไป 135 องศา ตามสมการที่ (32) ข้างล่างนี้

$$\begin{aligned}\psi_2(x,y) &= [\psi_h(x) + \mathrm{j}\,\psi_g(x)]\,\overline{[\psi_h(y) + \mathrm{j}\,\psi_g(y)]} \\ &= [\psi_h(x) + \mathrm{j}\,\psi_g(x)]\,[\psi_h(y) - \mathrm{j}\,\psi_g(y)] \\ &= \psi_h(x)\,\psi_h(y) + \psi_g(x)\,\psi_g(y) \\ &\quad + \mathrm{j}\,[\psi_g(x)\,\psi_h(y) - \psi_h(x)\,\psi_g(y)].\end{aligned} \tag{32}$$

หรือ

$$\psi_2(x,y) = \psi(x)\,\overline{\psi(y)}. \tag{33}$$

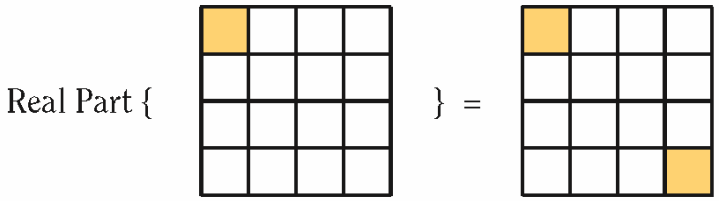
และ ไดอะแกรมของสเปกตรัมของผลคูณ ที่มีสเปกตรัมแบบข้างเดียว แสดงข้างล่างนี้



และโดยการคำนวณหาค่าจำนวนจริง ของสมการที่ (32) เราจะได้

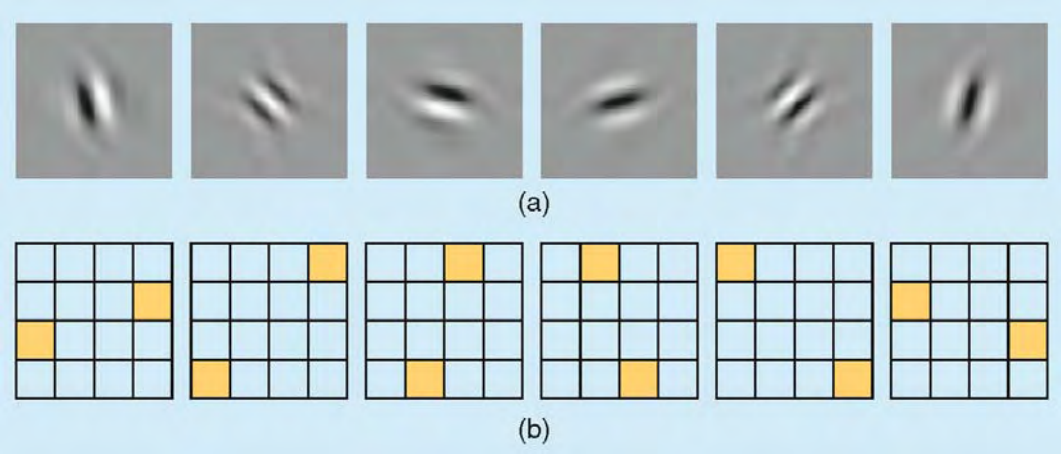
$$\text{Real Part}\{\psi_2(x,y)\} = \psi_h(x)\,\psi_h(y) + \psi_g(x)\,\psi_g(y), \tag{34}$$

และ ไดอะแกรมของสเปกตรัมของผลคูณของ HH ที่หันทิศ 135 องศา แสดงข้างล่างนี้



สำหรับเวฟเลตจำนวนจริง 4 ตัว ที่เหลือนั้น จะสามารถคำนวณจาก ผลเทนเซอร์ ของ $\phi(x)\psi(y), \psi(x)\phi(y), \phi(x)\bar{\psi}(y)$ และ $\psi(x)\bar{\phi}(y)$ โดยที่ $\phi(x) = \phi_h(x) + j\phi_g(x)$ และ $\psi(x) = \psi_h(x) + j\psi_g(x)$ ตามลำดับ

รูปที่ 4.2 แสดงเวฟเลตแบบมีทิศ จำนวน 6 ตัว เป็นที่น่าสังเกตว่า เวฟเลตทั้งหกตัว ไม่มีปัญหาสิ่งแปลกปน เหมือน เวฟเลตแบบแยกจากกันได้ เดิม รูปที่ 4.3 แสดงชุดคำสั่ง MATLAB ที่ใช้ในการคำนวณผลการแปลงเวฟเลตแบบมีทิศ



รูปที่ 4.2 Dual-tree wavelet จำนวนจริงสองมิติ รูป (a) แสดงเวฟเลตที่อยู่ในปริภูมิตำแหน่งของแถบความถี่ย่อย รูป (b) แสดงผลตอบสนองทางความถี่ฟูเรียร์

```
function w = dualtree2D(x, J, Faf, af)
% 2D Dual-Tree Discrete Wavelet Transform
%
% w = dualtree2D(x, J, Faf, af)
% INPUT:
% x - 2-D signal
% J - number of stages
% Faf - first stage filters
% af - filters for remaining stages
% OUPUT:
% w{i}{1:J+1}: tree i wavelet coeffs (i = 1,2)

[x1 w{1}{1}] = afb2D(x, Faf{1});
for k = 2:J
  [x1 w{k}{1}] = afb2D(x1, af{1});
end
w{J+1}{1} = x1;

[x2 w{1}{2}] = afb2D(x, Faf{2});
for k = 2:J
  [x2 w{k}{2}] = afb2D(x2, af{2});
end
w{J+1}{2} = x2;

for k = 1:J
  for m = 1:3
[w{k}{1}{m} w{k}{2}{m}] = pm(w{k}{1}{m},w{k}{2}{m});
  End
end
```

รูปที่ 4.3 ชุดคำสั่ง MATLAB สำหรับคำนวณผลการแปลงเวฟเลต dual-tree จำนวนจริง

5. ผลการทดลอง

ด้วยสมมติฐานที่ว่า เรามีข้อมูลเกี่ยวกับการเคลื่อนไหวเฟรมต่อเฟรมอย่างครบถ้วน ทำให้เราสามารถใช้ข้อมูลเหล่านี้ไป จัดสร้างเมทริกซ์ของตัวดำเนินการที่ใช้ในการสร้างชุดลำดับภาพความละเอียดต่ำจากภาพความละเอียดสูงยวดยิ่ง ดังสมการที่กำหนดไว้ใน (5) ในการทดลองของเรา พิกเซลของภาพที่นำมาทดสอบจะถูกเลื่อนอย่างสุ่มด้วยค่าที่เป็นเอกฐานจำนวนเต็ม ถูกทำให้ *พร่า* (blur) ด้วยฟังก์ชันแผ่แบบเกาส์ขนาด 4×4 ที่มีค่าเบี่ยงเบนเท่ากับ 1 และ *ถูกลดเวลาการชักตัวอย่าง* (downsampling) ลง 4 เท่า เพื่อสร้างเป็นชุดลำดับภาพความละเอียดต่ำจำนวน 16 ภาพ ต่อ หนึ่งภาพความละเอียดสูงต้นฉบับ โดยการเลือกภาพ 9 ภาพ จาก 16 ภาพ ดังกล่าวล่วงหน้า เราจะสามารถสร้างปริภูมิความละเอียดสูงยวดยิ่ง และ ภาพความละเอียดสูงยวดยิ่ง ได้ตามลำดับ จากนั้นเราจะทำการเปรียบเทียบ ระเบียบวิธีที่เรานำเสนอ กับ การรู้จำใบหน้าและเป้าหมายอัตโนมัติ ในระดับปริภูมิพิกเซล

5.1 ฐานข้อมูลที่ใช้ในการทดสอบ

ระเบียบวิธีปริภูมิใบหน้าเฉพาะความละเอียดสูงยวดยิ่งจะถูกนำมาใช้เป็นระเบียบวิธีตั้งต้นในการเปรียบเทียบสมรรถนะกับระเบียบวิธีที่เรานำเสนอ โดยทำการเปรียบเทียบด้วยฐานข้อมูลใบหน้าและเป้าหมายอัตโนมัติที่เป็นที่รู้จักกันดี ได้แก่ ฐานข้อมูลใบหน้า FERET Yale AR และ ORL (Phillips, 2000; Yale, 1997 & Martinez, 1998) และ ฐานข้อมูล MSTAR ซึ่งเป็นภาพ SAR ของยานพาหนะทางทหาร (Center, 1997) ตามลำดับ

5.1.1 ฐานข้อมูล FERET

โปรแกรมต่อต้านยาเสพติดของ DOD เป็นผู้สนับสนุนโปรแกรมทางเทคโนโลยีในการรู้จำใบหน้า (Facial Recognition Technology: FERET) และเป็นผู้พัฒนาฐานข้อมูล FERET โดยมี National Institute of Standards and Technology (NIST) เป็นผู้แจกจ่ายฐานข้อมูล ฐานข้อมูลนี้จัดเก็บในระหว่าง เดือนตุลาคม ค.ศ. 1993 ถึง สิงหาคม 1996 ฐานข้อมูลดังกล่าวนี้มีไว้เพื่อใช้ พัฒนา ทดสอบ และ ประเมิน ระเบียบวิธีที่ใช้ในการรู้จำใบหน้า

ภาพใบหน้าจะถูกจัดเก็บทั้งหมด 15 ครั้งย่อย แต่ละครั้งย่อยของการจัดเก็บจะใช้เวลาไม่เกิน 2 วัน เพื่อให้เกิดความสม่ำเสมอของฐานข้อมูล ภาพใบหน้าของแต่ละบุคคลจะถูกจัดเก็บโดยประมาณ 5 ถึง 11 ภาพถ่ายหน้าตรงสองภาพ (fa และ fb) ถูกจัดเก็บ โดยในภาพถ่ายที่จะสองจะถูกขอร้องให้แสดงอารมณ์แตกต่างกันออกไปจากหน้าแรก ภาพถ่ายหน้าตรงภาพที่สามจะถูกถ่ายโดยใช้กล้องถ่ายภาพกล้องอื่น และมีแสงเงาที่แตกต่างกันออกไป (ภาพที่สามนี้

จะถูกอ้างอิงว่าคือภาพ fc) ภาพใบหน้าอื่นจะถูกจัดเก็บโดยเห็นทางด้านข้าง ซ้าย และ ขวา และ ในบางภาพอาจถูกขอให้ใส่หมวก หรือ แว่นตา เพื่อเป็นข้อมูลภาพในชุดที่สอง ในบางครั้งบุคคลอาจถูกขอให้ถ่ายอีกครั้งในวันรุ่งขึ้น เพื่อเป็นชุดภาพที่ทำซ้ำ ที่ซึ่งชุดภาพที่ทำซ้ำนี้จะมีความแตกต่างกันทางขยาด มุมมอง อารมณ์ และ แสงเงาที่ปรากฏบนใบหน้า

ฐานข้อมูล ชุด แรกประกอบด้วย ภาพใบหน้าทั้งหมด 14,126 ภาพ ของบุคคลทั้งหมด 1,199 คน และ 365 ชุดภาพที่ทำซ้ำ สำหรับภาพบุคคลบางคนอาจถ่ายครั้งแรก และ ครั้งที่สองนานต่างกันถึง 2 ปี ชุดของภาพบุคคลบุคคลหนึ่งแสดงดังในรูปที่ 5.1

ด้วยเหตุที่ฐานข้อมูลนี้มีภาพที่มีความหลากหลายมาก จึงเหมาะสมที่จะใช้กับการทดสอบสมมรถนะของการรู้จำใบหน้า

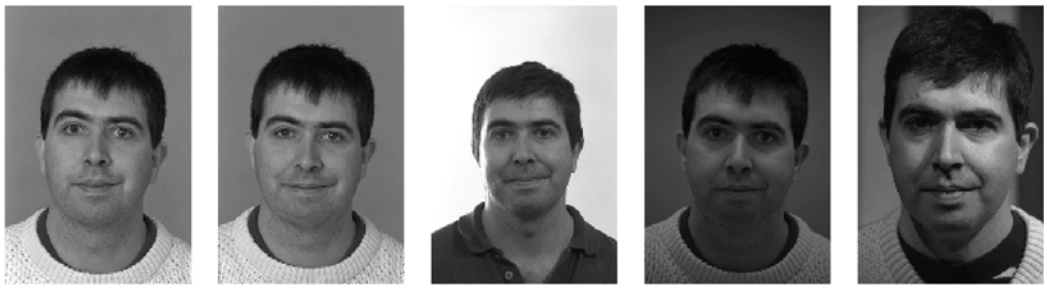
5.1.2 ฐานข้อมูล Yale

ฐานข้อมูล Yale ประกอบด้วยภาพบุคคลจำนวน 165 ภาพ ที่เก็บจากบุคคล 15 คน หนึ่งบุคคลมี 11 ภาพ สำหรับแต่ละบุคคลจะประกอบด้วยภาพที่แสดงความรู้สึกทางใบหน้า เช่น แสงตรง ใส่แว่นตา มีความสุข แสงส่องจากทางซ้าย ไม่ใส่แว่นตา ปกติ แสงส่องจากทางขวา เศร้า ง่วง ประหลาด ใจ และ ทำหน้าหย่น รูปตัวอย่างทั้งหมดของบุคคลหนึ่งจากฐานข้อมูล Yale แสดงในรูปที่ 5.2 โดยที่ภาพแต่ละภาพได้ถูกทำการครอบตัดและปรับให้มีขนาด 100×80 พิกเซล

ในการทดลองทั้งหมด ภาพตัวอย่างห้าภาพแรก (แสงตรง ใส่แว่นตา มีความสุข แสงส่องจากทางซ้าย และ ไม่ใส่แว่นตา) จะใช้สำหรับการฝึกฝน และหกภาพที่เหลือ (ปกติ แสงส่องจากทางขวา เศร้า ง่วง ประหลาด ใจ และ ทำหน้าหย่น) จะใช้เพื่อการทดสอบหาความถูกต้องในการรู้จำ

5.1.3 ฐานข้อมูล AR

ฐานข้อมูลหน้า AR ถูกสร้างโดย Robert Benavente และ Aleix Martinez วิศวกรประจำศูนย์คอมพิวเตอร์วิทัศน์ (CVC) ที่ U.A.B. ฐานข้อมูลนี้ประกอบด้วยภาพสีมากกว่า 4000 ภาพ ของภาพใบหน้าของบุคคลจำนวน 126 คน โดยในจำนวนนี้เป็นผู้ชาย 70 คนและผู้หญิง 56 คน ภาพใบหน้าเป็นภาพใบหน้าตรงที่มีการแสดงออกทางใบหน้าต่างๆกัน มีการส่องไฟที่ต่างกัน และ มี *การบดบัง* (occlusions) การบดบังที่ว่า ได้แก่ การใส่แว่นกันแดด และ ผ้าพันคอ ภาพทั้งหมดถูกถ่ายที่ CVC โดยการควบคุมสภาวะอย่างเคร่งครัด แต่ไม่มีข้อห้ามในการเครื่องแต่งกาย (เสื้อผ้า แว่นตา หรือ อื่นๆ) เครื่องสำอาง ทรงผม หรือ อื่นๆ บุคคลแต่ละคนจะถูกเรียกเข้าไปถ่ายภาพเก็บไว้จำนวน 2 ช่วง แต่ละช่วงคั่นไว้สองสัปดาห์ (14 วัน) และถ่ายภาพในลักษณะเดียวกันทั้งสองช่วง



รูปที่ 5.1 ภาพตัวอย่างของบุคคลหนึ่งในฐานข้อมูล FERET สำหรับบุคคลนี้ ชุดที่ถ่ายซ้ำที่ 1 (duplicate I) จะถ่ายหลัง fa หนึ่งปี เช่นเดียวกับ ชุดที่ถ่ายซ้ำชุอที่ 2 และ ภาพ fb

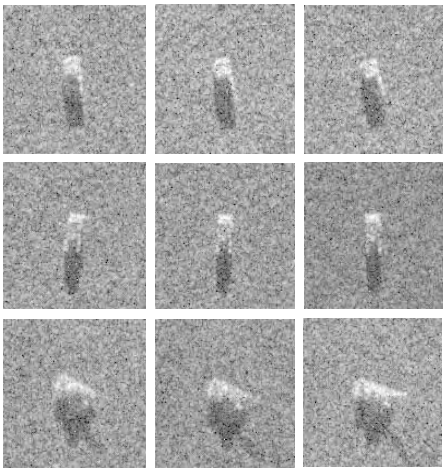
รูปที่ 5.2 ภาพตัวอย่างของบุคคลหนึ่งในฐานข้อมูล Yale

รูปที่ 5.3 ภาพตัวอย่างของบุคคลหนึ่งในฐานข้อมูล AR

ในการทดลอง เราจะเลือกภาพที่ไม่มีการบดบัง (การใส่แว่นกันแดดและผ้าพันคอ) จำนวน 14 ของแต่ละหัวบุคคล ดังที่แสดงในรูปที่ 5.3 โดยที่ภาพแต่ละภาพได้ถูกทำการครอบตัดและปรับให้มีขนาด 112×92 พิกเซล จากนั้นแปลงภาพสีเหล่านี้ให้เป็นภาพระดับเทา ภาพตัวอย่างห้าภาพแรกจะใช้ในการฝึกฝนและภาพที่เหลือจะใช้เพื่อการทดสอบผลของการรู้จำ

5.1.4 ฐานข้อมูล ORL

ฐานข้อมูล ORL ประกอบด้วยภาพของบุคคลจำนวน 40 คน แต่ละคนมีภาพอยู่ 10 ภาพ ภาพเหล่านั้นถูกเก็บไว้ในเวลาที่แตกต่างกัน มีแสงเงาที่แตกต่างกัน มีอารมณ์ที่แตกต่างกัน (เปิด/ปิดตา ยิ้ม/ไม่ยิ้ม) และมีรายละเอียดอื่นๆ (ใส่แว่น/ไม่ใส่แว่น) ภาพทั้งหมดถ่ายโดยมีฉากหลังดำสนิทเป็นเนื้อเดียวกัน โดยถ่ายในระหว่างยืน หน้าตรง (อาจมีการหันข้างเล็กน้อย ถึงประมาณ 20 องศา)



รูปที่ 5.4 ตัวอย่างรูปภาพ SAR ของฐานข้อมูล MSTAR: แถวด้านบนคือ BMP2 APCs
แถวตรงกลางคือ BTR70 APCs และแถวล่างเป็นถัง T72

5.1.5 ฐานข้อมูล MSTAR

ฐานข้อมูล MSTAR เป็นฐานข้อมูลสำหรับเผยแผ่สู่สาธารณะ ประกอบด้วยข้อมูลภาพเรดาร์ช่องเปิดเชิงสังเคราะห์ความละเอียดสูง (SAR) ซึ่งถูกรวบรวมโดยหน่วยงาน DARPA/ห้องปฏิบัติการ Wright ซึ่งเป็นโครงการในการตรวจหาที่ตั้งและการรู้จำเป้าหมายเคลื่อนไหวและอยู่กับที่ (Moving and Stationary Target Acquisition and Recognition: MSTAR) ข้อมูลประกอบด้วยภาพ SAR ขนาด 128×128 ของยานพาหนะทหารสามชนิด ได้แก่ รถหุ้มเกราะส่วนบุคคลผู้ (APCs) รุ่น BMP2, APCs รุ่น BTR70 และ รถถัง รุ่น T72 ภาพเป้าหมายตัวอย่างจากฐานข้อมูล MSTAR แสดงดังในรูปที่ 5.4 เนื่องจากฐานข้อมูล MSTAR มีขนาดใหญ่ ภาพทั้งหมดได้ถูกครอบตัดจากส่วนกลางให้มีขนาด 32×32 พิกเซล เพื่อความรวดเร็วในการทดสอบ

ตารางที่ 5.1 ภาพ MSTAR ที่ประกอบด้วยชุดข้อมูลสำหรับการฝึกฝน

	Vehicle No.	Serial No.	Depression Angle	Images
BMP-2	1	9563	17^{o}	233
	2	9566		231
	3	C21		233
BTR-70	1	C71	17^{o}	233
T-72	1	132	17^{o}	232
	2	812		231
	3	S7		228

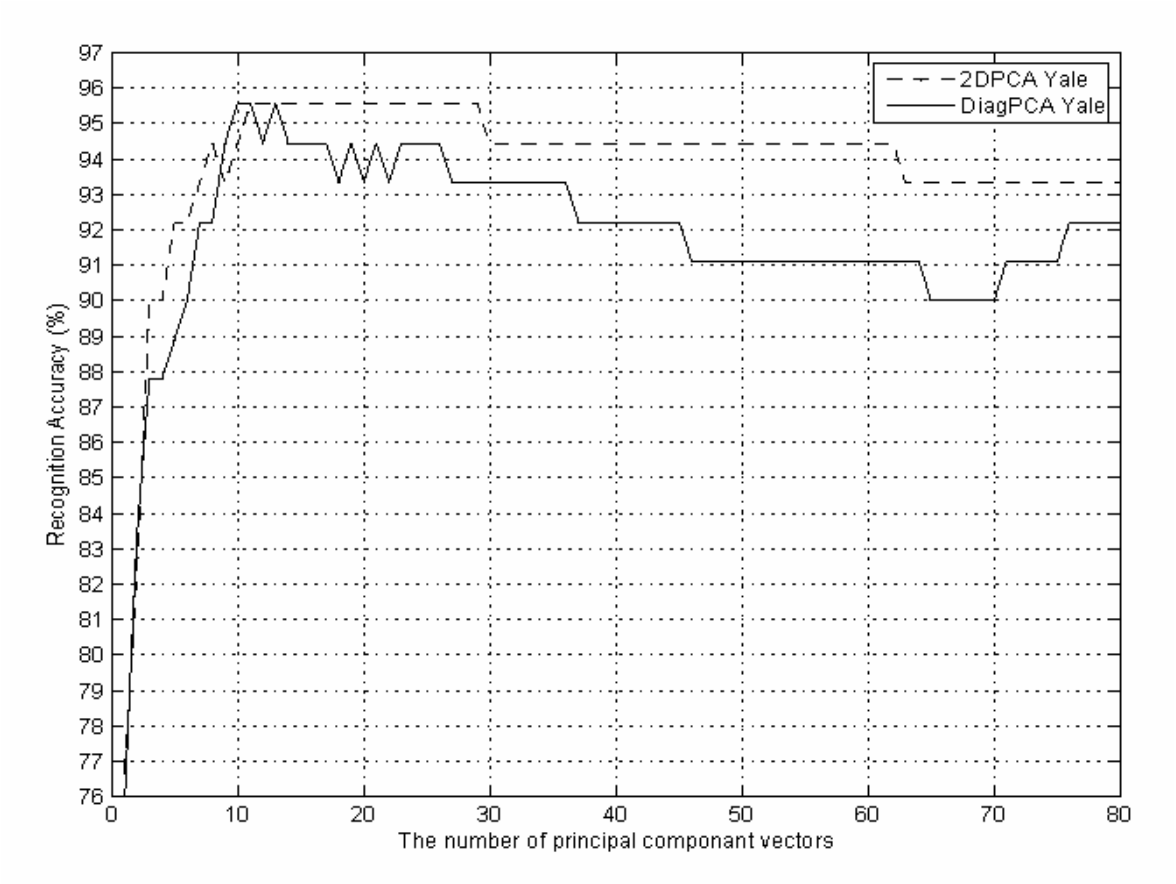
ตารางที่ 5.2 ภาพ MSTAR ที่ประกอบด้วยชุดข้อมูลสำหรับการทดสอบ

	Vehicle No.	Serial No.	Depression Angle	Images
BMP-2	1	9563	15^{o}	195
	2	9566		196
	3	C21		196
BTR-70	1	C71	15^{o}	196
T-72	1	132	15^{o}	196
	2	812		195
	3	S7		191

ตารางที่ 5.1 และ 5.2 แสดงรายละเอียดของภาพเป้าหมายที่ใช้ในการฝึกฝนและการทดสอบ โดยที่ *มุมกัม* (depression) หมายถึงมุมที่บีมของเสาอากาศมองมาเป้าหมายจากด้านข้างของเครื่องบิน อนึ่งภาพมุมกัม SAR ถูกจัดเก็บในช่วงเวลาที่แตกต่างกัน ข้อมูลชุดทดสอบที่มุมกัมที่ต่างจากชุดข้อมูลที่ใช้ในการฝึกฝน ถูกนำมาใช้ในการทดสอบประสิทธิภาพ ข้อมูลชุดทดสอบนี้ จึงสามารถนำมาใช้เป็นตัวแทนของข้อมูลทดสอบโดยทั่วไปได้อย่างเหมาะสม

5.2 ผลการทดลอง 2DPCA ระบุทิศทาง

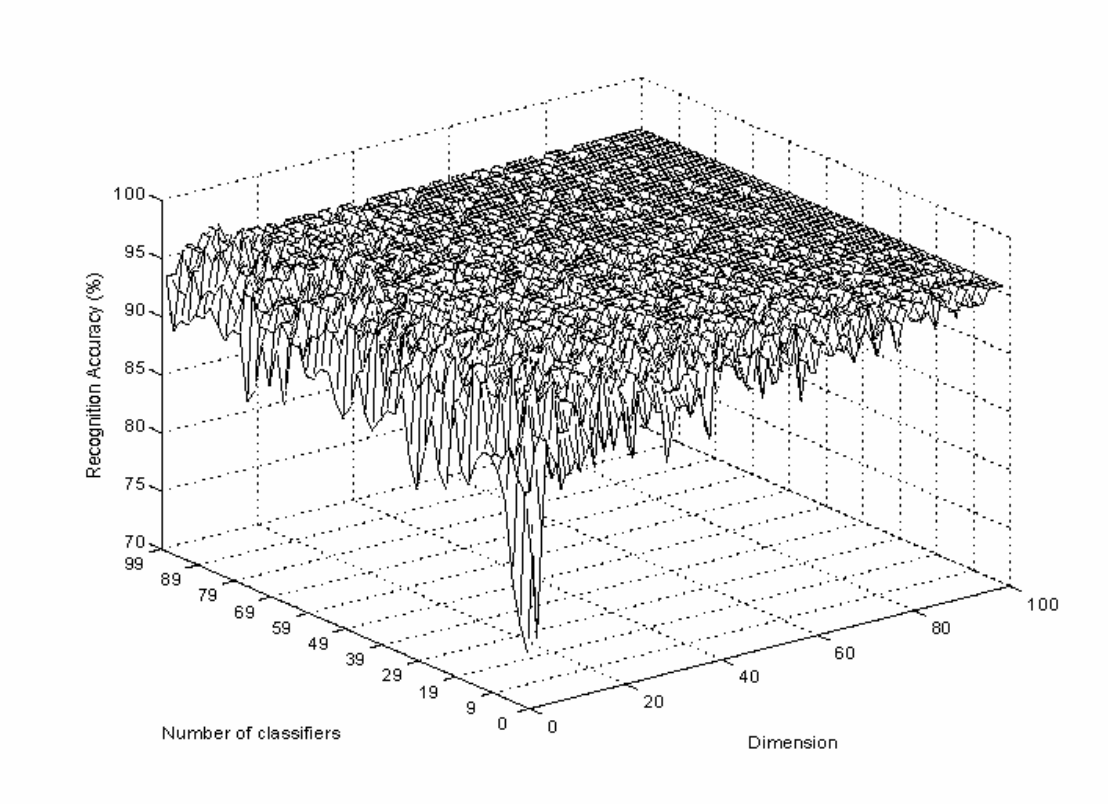
บทย่อยนี้จะกล่าวถึงผลการทดลอง 2DPCA แนวขวาง ซึ่งเป็นพื้นฐานของ 2DPCA ระบุทิศทาง โดยพัฒนาตามแนวคิดระบบจำแนกแบบหลายตัวด้วยวิธีปริภูมิย่อยสุ่ม ซึ่งได้แสดงชุดคำสั่งเทียม ในภาคผนวก ข. แล้ว โดยเราจะใช้ 2DPCA ปกติ เป็นสมรรถนะพื้นฐานในการเปรียบเทียบ และทำการทดสอบบนฐานข้อมูล Yale



รูปที่ 5.5 สมรรถนะการรู้จำใบหน้าระหว่าง 2DPCA แนวระนาบราบ กับ แนวขวางบนฐานข้อมูล Yale

เบื้องต้น ในการทดลองแรก เราจะเปรียบเทียบสมรรถนะของ 2DPCA แนวระนาบ กับ 2DPCA แนวขวาง โดยมีการเปลี่ยนจำนวนของ *เวกเตอร์ส่วนประกอบมุขสำคัญ* (principal component vector) ผลการทดลองแสดงดังในรูปที่ 5.5 2DPCA แนวขวางให้สมรรถนะที่ดีกว่า 2DPCA แนวระนาบที่จำนวนเวกเตอร์ที่น้อยๆ (เป็นการทดลอง สมมติฐานแรกที่เกี่ยวกับความเป็นไปได้ทางชีวภาพแห่งการมองเห็น ที่ว่า ข่าวสารที่สำคัญจะถูกเก็บไว้ในบางทิศทาง มากว่า ทิศอื่น)

การทดลองที่สอง เป็นการทดสอบสมมติฐานที่ว่าปมประสาทของการมองเห็น ช่วยทำงานกันแบบไม่เป็นระเบียบ โดยกรใช้วิธีปริภูมิย่อยสุ่ม ในการช่วยการรู้จำแบบหลายการตัดสินใจ ในที่นี้เราจะทดลองโดยการเปลี่ยนให้จำนวน ตัวจำแนกในคณะกรรมการตัดสินใจ มี จำนวนเป็นเลขคี่ตั้งแต่ 1 ถึง 99 ที่เราเลือกจำนวนเป็นเลขคี่ ก็เพื่อให้การตัดสินใจจากการลงคะแนนเสียงไม่กำกวม ขนาดมิติ ของ ปริภูมิย่อยสุ่ม (random subspace) จะมีขนาดเท่ากับ 1 ถึง 100 (100 คือ ความสูงของภาพในที่นี้) ถ้าขนาดมิติเท่ากับ 100 เราจะได้ 2DPCA แนวขวาง ปกติ 1 ตัวจำแนก รูปที่ 5.6 แสดงความแม่นยำในการรู้จำของ 2DPCA แนวขวางกับปริภูมิย่อยสุ่มที่พารามิเตอร์ต่างๆ ค่าความแม่นยำสูงสุดของระเบียบวิธีทั้ง 3 ระเบียบวิธี ได้แสดงไว้ใน ตารางที่ 5.3

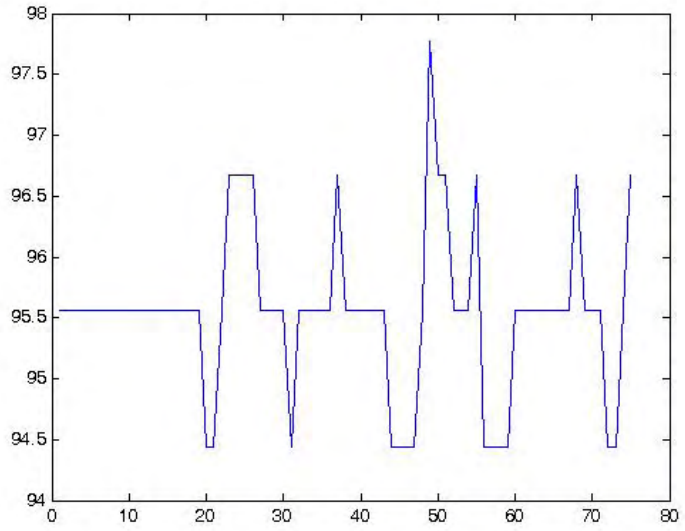


รูปที่ 5.6 ความแม่นยำในการรู้จำของ 2DPCA แนวขวางกับวิธีปริภูมิย่อยสุ่ม บนฐานข้อมูล Yale

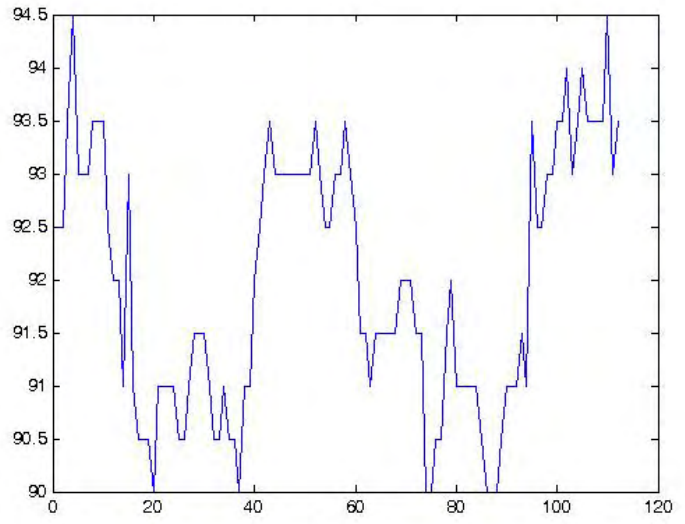
ตารางที่ 5.3 ตารางเปรียบเทียบสมรรถนะ 2DPCA ระนาบราบ 2DPCA แนวขวาง และ 2DPCA แนวขวางด้วยวิธีปริภูมิย่อยสุ่ม บนฐานข้อมูล Yale

Database	*ระเบียบวิธี*	*ความแม่นยำ*	*จำนวนเวกเตอร์เฉพาะ*	*มิติที่ใช้*	*จำนวนตัวจำแนก*
Yale	2DPCA	95.56%	11	100	1
	2DPCA แบบขวาง	95.56%	10	100	1
	RSM+2DPCA แบบขวาง	97.78%	10	4	9

จะเห็นได้ว่า RSM+2DPCA แบบขวางให้ความแม่นยำในการรู้จำได้ดีที่สุด จึงน่าจะมีความเป็นไปได้ทางชีวภาพแห่งการมองเห็นมากที่สุด



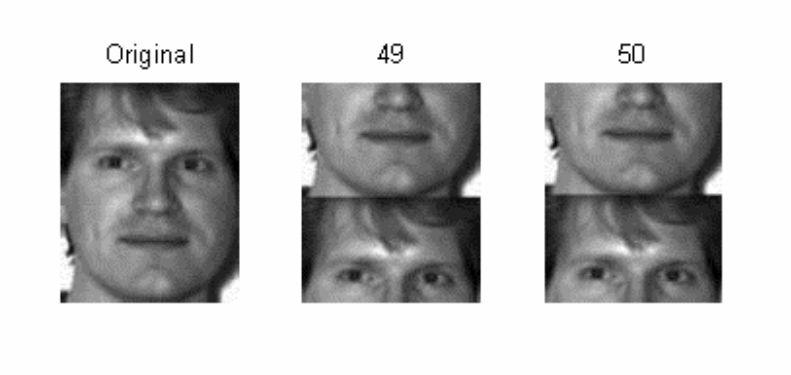
รูปที่ 5.8 สมรรถนะการรู้จำใบหน้าบนฐานข้อมูล Yale ที่ทดสอบกับ 2DPCA ไขว้ที่ i^{th}



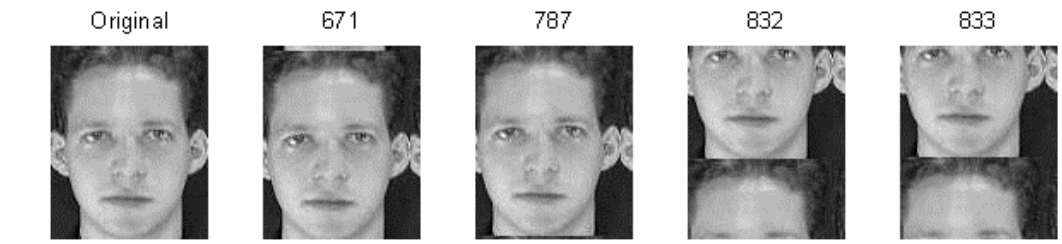
รูปที่ 5.9 สมรรถนะการรู้จำใบหน้าบนฐานข้อมูล ORL ที่ทดสอบกับ 2DPCA ไขว้ที่ i^{th}

5.3 ผลการทดลอง 2DPCA ไขว้

บทย่อยนี้จะกล่าวถึงผลการทดลอง 2DPCA ไขว้ ซึ่งเป็นหนึ่งใน 2DPCA ระบุทิศทาง ซึ่งได้แสดงชุดคำสั่งเทียม ในภาคผนวก ข. แล้ว โดยเราได้ทดลองกับฐานข้อมูลใบหน้า Yale และ ORL โดยมีความแม่นยำในการรู้จำสูงสุดเท่ากับ 97.5 และ 94.5 เปอร์เซ็นต์ ตามลำดับ รูปที่ 5.8 และ 5.9 แสดงสมรรถนะของ 2DPCA ไขว้ บนฐานข้อมูล Yale และ ORL ตามลำดับ เป็นที่น่าสังเกตว่า ที่ i^{th} เท่ากับ คือ 2DPCA แนวราบ ที่ไม่มีการหมุนแบบเป็นวงกลมเพื่อหาเมทริกซ์ความแปรปรวนร่วมเกี่ยวไขว้



รูปที่ 5.9 ภาพใบหน้าที่ถูกแปลงไขว้ ที่ให้ความแม่นยำใน
การรู้จำที่ดีที่สุด ของฐานข้อมูล Yale



รูปที่ 5.10 ภาพใบหน้าที่ถูกแปลงไขว้ ที่ให้ความแม่นยำใน
การรู้จำที่ดีที่สุด ของฐานข้อมูล ORL

จะเห็นว่า เมื่อเปรียบเทียบกับการทดลองที่ 5.2 แล้ว 2DPCA ไขว้ มีความแม่นยำในการรู้จำที่ดีที่สุดดีกว่า 2DPCA แนวราบ และ 2DPCA ระบุทิศทาง (มีความแม่นยำในการรู้จำที่ดีที่สุดเท่ากับ 95.56% เท่ากัน) หรือ ดีกว่า แนวพื้นฐาน (base line) ที่เป็น 2DPCA แนวราบ อยู่มากกว่า 2 เปอร์เซ็นต์ (ดูรูปที่ 5.8 และ 5.9) รูปที่ 5.10 และ 5.11 แสดงรูปในฐานข้อมูล Yale และ ORL ที่เกิดการแปลงเลื่อน ที่ให้ความแม่นยำในการรู้จำสูงสุด จากรูปที่ 5.10 และ 5.11 เราอาจอนุมานได้ว่าสหสัมพันธ์ระหว่างตำแหน่งของหู กับ ตา หรือ ตา กับ ไรผม อาจมีผลต่ออำนาจในการจำแนกภาพวัตถุ

แม้ว่า ตัวบ่งต่างที่สกัดได้จาก 2DPCA ไขว้ จะให้ความแม่นยำในการรู้จำสงสุด แต่ในตัวบ่งต่างไขว้ บางตัวจะให้ผลในการรู้จำที่ต่ำกว่ามาตรฐาน ซึ่งก็สมเหตุผลเพราะในบางทิศเราอาจได้ผลรวมของตัวบ่งต่างที่ควบรวม ตัวบ่งต่างที่มีอำนาจจำแนกเข้าด้วยกัน ทำให้ตัวบ่งต่างที่เลหือมีอำนาจในการจำแนกไม่ดีพอ

นอกจากนั้น 2DPCA ไขว้ ซึ่งในที่นี้เป็น ตัวบ่งต่างที่นำเสนอเพื่อใช้ในการสร้าง ตัวจำแนกอย่างเข้มแข็ง เพียงตัวเดียว ยังมีประสิทธิภาพด้อยกว่า RSM+2DPCA แบบขวาง ซึ่งเป็น ระบบจำแนกแบบหลายตัว โดยที่ตัวจำแนกแต่ละตัวเป็น ตัวจำแนกอย่างอ่อนแอ แต่ทำการตัดสินใจร่วมกันเป็นคณะ โดยการลงคะแนนเสียง (สมรรถนะของ 2DPCA ไขว้ และ RSM+2DPCA แบบขวาง เท่ากับ 97.78 และ 95.5 เปอร์เซ็นต์ ตามลำดับ)

เราจึงสนใจ พัฒนา 2DPCA ไขว้ ให้มีการตัดสินใจแบบคณะ ให้ตัวจำแนกแต่ละตัว รับตัวบ่งต่างจาก 2DPCA ไขว้ที่ต่างกัน และ นำผลของการตัดสินใจมาลงคะแนนเสียง เพื่อให้ได้ผลการตัดสินใจในขั้นสุดท้าย ซึ่งในการพัฒนา เราจะได้ผลการทดลองที่ดีขึ้น ดังรายละเอียดต่อไปนี้

- ☐ On Yale database ใช้ภาพในการฝึกฝน 5 ภาพ และทดสอบ 6 ภาพ
  - ■ Max rate of 2DPCA = 95.56%
  - ■ Max rate of proposed method
    = 97.78% @ 49th Shifting
- ☐ On ORL database ใช้ภาพในการฝึกฝน 5 ภาพ และทดสอบ 6 ภาพ
  - ■ Max rate of 2DPCA = 92.5%
  - ■ Max rate of proposed method
    = 94.5% @ 4th and 110th Shifting

ตารางที่ 5.4 แสดงการเปรียบเทียบ สมรรถนะของ 2DPCA แนวราบ 2DPCA ไขว้ ตัวจำแนกเดียว และ 2DPCA ไขว้ หลายตัวจำแนก เพื่อหลีกเลี่ยงความสับสน เราจะเรียก 2DPCA ไขว้ ตัวจำแนกเดียว และ 2DPCA ไขว้ หลายตัวจำแนก สั้นๆว่า 2DPCA ไขว้ และ MCS-2DPCA ไขว้ ตามลำดับ ในตารางที่ 5.4 *d* หมายถึง จำนวนเวกเตอร์เฉพาะที่ 2DPCA ให้ผลในการรู้จำดีที่สุด และ *L* หมายถึง จำนวน 2DPCA ไขว้ ที่เกิดจากเลื่อนภาพต้นแบบเป็นเมทริกซ์ภาพ B จำนวนทั้งหมด *L* ครั้ง

จะเห็นได้ว่า MCS-2DPCA ไขว้ ที่นำเสนอสามารถเพิ่มประสิทธิภาพในการรู้จำให้ดีขึ้นได้ อย่างไรก็ดีเราใช้ 2DPCA ไขว้ ที่มีการเลื่อนอย่างต่อเนื่อง ซึ่งในบางตัวบ่งต่าง อาจจะมีความซ้ำซ้อน หรือ มีอำนาจในการตัดสินใจต่ำ ดังนั้น เราจึงควรเลือกเฉพาะ 2DPCA ไขว้ เฉพาะที่มีอำนาจในการตัดสินใจสูง

ในความเป็นจริง เราสามารถสร้าง ชุดของ 2DPCA ไขว้ ได้อีกมากมายหลายชุด เช่น จาก 2DPCA แนวดิ่ง หรือ 2DPCA แนวขวางของแนวราบ หรือ 2DPCA แนวขวางของแนวดิ่ง เป็นที่น่าสังเกตว่า 2DPCA ใช้วิธีบวกรวมกำลังงานของสหสัมพันธ์ในแนวแกนที่ไม่ได้ถูกเลือก และคงตัวบ่งต่างที่มีอำนาจไว้ในแนวแกนที่ถูกเลือก อย่างไรก็ดี เป็นการยากที่จะรูว่าแนวแกนไหน เป็นแนวแกนที่ควรคัดเลือก 2DPCA และส่วนขยาย ใช้แนวความคิดในการบวกรวมในแนวแกนข้อมูลภาพที่ไม่ได้เลือก เพื่อหลีกเลี่ยงปัญหา คำสาปมิติ เป็นที่เข้าใจได้ว่า สิ่งมีชีวิต มีการเลือกทิศที่ทำให้อำนาจการรู้จำดีขึ้น โดยผ่านการเรียนรู้ เช่น ข่ายวงจรประสาท โดยอาศัยวิวัฒนาการเป็นเวลายาวนานในการเลือกสรรทิศที่ของตัวสกัดลักษณะบ่งต่างสามารถทำงานได้อย่างเหมาะสมที่สุด ซึ่งอาจจะเป็นทิศที่ขึ้นอาจอยู่กับสิ่งมีชีวิตแต่ละชีวิตก็ได้ ซึ่งพบได้ในเซลของจอประสาทตา และ ส่วนรับรู้ในการมองเห็น ในสัตว์เลี้ยงลูกด้วยนมส่วนใหญ่

ตารางที่ 5.4 ตารางเปรียบเทียบสมรรถนะ 2DPCA ระนาบราบ 2DPCA ไขว้ และ 2DPCA ไขว้ แบบหลายตัวจำแนก บนฐานข้อมูล Yale และ ORL

Database	*ระเบียบวิธี*	*ความแม่นยำ (%)*	*d*	*L*
Yale	2DPCA	95.56	20	1
	2DPCA ไขว้	97.5	20	1
	MCS-2DPCA ไขว้	97.78	20	49
ORL	2DPCA	92.5	5	1
	2DPCA ไขว้	94.5	5	1
	MCS-2DPCA ไขว้	95	5	671

หากเราขยายความคิดนี้ ออกไปซักเล็กน้อย เราอาจจะเห็นว่า ความเข้มของอำนาจจำแนกในแต่ละแกนของภาพ ก็ไม่น่าจะเท่ากัน ซึ่งในเซลประสาทของเราที่ผ่านวิวัฒนาการมาอย่างยาวนานได้ทำหน้าที่หาฟังก์ชันในการถ่วงน้ำหนักที่เหมาะสมแล้ว ได้ฟังก์ชันฐานหลักที่ดีในเกือบทุกเงื่อนไข ในทางคณิตศาสตร์ ฟังก์ชันเวฟเลต เป็น ฟังก์ชัน อีกตัวหนึ่ง ที่ดีที่สุดในการแทนรูปอย่างไม่มีเงื่อนไข (unconditional bases) อีกทางเลือกหนึ่งของการสร้างเมทริกซ์ความแปรปรวนร่วมเกี่ยวที่มีการลดขนาดมิติ เราอาจใช้ผลการแปลงเวฟเลต กับเมทริกซ์ความแปรปรวนร่วมเกี่ยวของ PCA เพื่อสร้างเมทริกซ์ความแปรปรวนร่วมเกี่ยวที่มีขนาดมิติเล็กลง และ สามารถระบุทิศได้

เร็วๆนี้ *ผลการแปลงกาบอ* (Gabor transform) ได้ถูกนำมาใช้เป็นการประมวลผลส่วนหน้าให้กับ การหา PCA เนื่องด้วยว่า ฟิวเตอร์กาบอ (Gabor filter) นั้นคงทนต่อความสว่าง และ คอนทราส (contrast) รวมไปถึงการหมุนทั้งในแนวแกนภาพสองมิติ และ สามมิติ ในลักษณะเดียวกันนี้ เราอาจลดความแปรเปลี่ยนของเมทริกซ์ความแปรปรวนร่วมเกี่ยวหนึ่งมิติ โดยอาศัยการป้อน สมาชิกของเมทริกซ์ความแปรปรวนร่วมเกี่ยวหนึ่งมิติ ผ่านชุดตัวกรองแบบหลายระดับความละเอียด (multiresolution filter banks) หรือ ผลการแปลงเวฟเลตแบบมีทิศด้วยวิธีนี้ เราอาจเปลี่ยนวิธีในการสร้างเมทริกซ์ความแปรปรวนร่วมเกี่ยวของ 2DPCA จาก PCA สำหรับ การรู้จำภาพใบหน้า และ เป้าหมายอัตโนมัติ

โดยอาศัย การสแกนแบบ Hilbert หรือ บริเวณที่มีกิจกรรมสูง เราสามารถเพิ่มความสามารถในการคำนวณ PCA และ 2DPCA แบบมีทิศให้เร็วและเหมาะสมมากขึ้นได้

5.4 ผลการทดลองการสร้างคืนปริภูมิความละเอียดสูงยวดยิ่งหนึ่งมิติแบบคลาสเฉพาะ

บทย่อยนี้จะกล่าวถึง สมรรถนะในการรู้จำภาพวัตถุโดยการสร้างคืนความละเอียดสูงยวดยิ่งในระดับปริภูมิ เปรียบเทียบ กับ การรู้จำภาพวัตถุโดยการสร้างคืนความละเอียดสูงยวดยิ่งในระดับพิกเซล โดยใช้การรู้จำภาพของวัตถุที่ภาพรายละเอียดต่ำหนึ่งภาพเป็นแนวฐาน (baseline) ในที่นี้เราทดลองกับฐานข้อมูลภาพใบหน้า Yale และ AR และ ภาพเห้าหมายที่เป้ฯยานพาหนะทางทหารบนฐานข้อมูล MSTAR ความแตกต่างระหว่างงานวิจัยนี้และงานวิจัยอื่น อยู่ที่เราสนใจเวกเตอร์เฉพาะที่ดำรงคุณสมบัติของอำนาจจำแนกด้วย มิใช่เพียงแต่อำนาจในการแทนรูปของ PCA เท่านั้น

ในงานวิจัยนี้ เราใช้การรู้จำแบบ *คลาสเฉพาะ* (class specific) ที่เรียกว่า คลาสเฉพาะนั้นหมายถึงว่าเราแยกตัวอย่างออกตามคลาสที่ทราบ และในแต่ละคลาส เราจะหาเวกเตอร์เฉพาะของแต่ละคลาส ดังนั้น หากเรามี จำนวนคลาส เท่ากับ C เราจะมีเวกเตอร์เฉพาะจำนวน C ชุด เมื่อเราได้เวกเตอร์เฉพาะจำนวน C ชุดแล้ว ในการรู้จำ เราจะนำภาพวัตถุที่ต้องการทดสอบ (ไม่ทราบคลาส ต้องการหาคลาสให้กับภาพวัตถุนี้) โดยทำการฉายลงบนเวกเตอร์เฉพาะของแต่ละคลาส และทำการสร้างคืนภาพ โดยการหาผลรวมถ่วงน้ำหนักของเวกเตอร์เฉพาะของแต่ละคลาส เราจะได้ภาพที่สร้างคืนมาทั้งหมด C ภาพ จากนั้นเรานำภาพทั้งหมดมาหาระยะห่าง (distance) กับ ภาพทดสอบเดิม คลาสใดที่ให้ระยะห่างระหว่างภาพน้อยที่สุด คือ คลาสที่เราจะตัดสินใจ

โดยในที่นี้เราประยุกต์ใช้ การรู้จำแบบ คลาสเฉพาะ เมื่อมีการสร้างคืนภาพความละเอียดสูงยวดยิ่งในระดับพิกเซล (ขนาดภาพ) หรือ เมื่อมีการสร้างคืนความละเอียดสูงยวดยิ่งในระดับปริภูมิย่อย แล้วเปรียบเทียบสมรรถนะที่ได้ ตารางที่ 5.5 และ 5.6 แสดงความสามารถในการรู้จำของระเบียบวิธีที่กล่าวถึง ตามลำดับ จะเห็นว่าความสามารถในการรู้จำของการสร้างคืนในระดับพิกเซลสูงกว่าในระดับปริภูมิย่อย แต่ข้อได้เปรียบของแบบปริภูมิย่อยคือ ความรวดเร็วในเฟสของการตัดสินใจ ตารางที่ 5.7 ตารางเปรียบเทียบสมรรถนะ LR PCA PCA สร้างคืนระดับพิกเซล และ PCA สร้างคืนระดับปริภูมิย่อย เช่นกัน จะเห็นว่าความสามารถในการรู้จำของการสร้างคืนในระดับพิกเซลสูงกว่าในระดับปริภูมิย่อย แต่ข้อได้เปรียบของแบบปริภูมิย่อยคือ ความรวดเร็วในเฟสของการตัดสินใจ ในที่นี้ LR PCA หมายถึง การรู้จำโดยใช้เวกเตอร์เฉพาะในระดับความละเอียดต่ำ รูปที่ 5.11 และ 5.12 แสดงการสร้างคืนของภาพตัวอย่าง ในระดับ พิกเซล และ ปริภูมิย่อย ตามลำดับ

ตารางที่ 5.5 ความสามารถในการรู้จำ ด้วยการสร้างคืนในระดับปริภูมิพิกเซล

	BMP-2	BTR-70	T-72	Percent
BMP-2	526	0	61	89.61
BTR-70	103	0	93	0
T-72	19	0	563	96.74
Avg. -	-	-	-	79.78

ตารางที่ 5.6 ความสามารถในการรู้จำ ด้วยการสร้างคืนในระดับปริภูมิย่อย

	BMP-2	BTR-70	T-72	Percent
BMP-2	526	0	61	89.61
BTR-70	116	0	80	0
T-72	41	0	541	92.96
Avg. -	-	-	-	78.17

ตารางที่ 5.7 ตารางเปรียบเทียบสมรรถนะ LR PCA PCA สร้างคืนระดับพิกเซล และ PCA สร้างคืนระดับปริภูมิย่อย

Database Method	CSS-PCA	Pixel-Domain SR-CSS	Eigen-Domain SR-CSS
Yale	88.89	88.56	81.11
AR	88.00	88.00	87.50
MSTAR	77.44	79.78	78.17

รูปที่ 5.11 ภาพตัวอย่างของการสร้างคืนความละเอียดสูงยวดยิ่งภาพในระดับพิกเซล

รูปที่ 5.12 ภาพตัวอย่างของการสร้างคืนความละเอียดสูงยวดยิ่งภาพในระดับเวกเตอร์เฉพาะ

6. สรุปและข้อเสนอแนะ

6.1 สรุป

งานวิจัยนี้ได้ทำการทดลองการรู้จำใบหน้าและเป้าหมายอัตโนมัติ โดยเน้นการสร้างคืนความละเอียดสูงยวดยิ่งในระดับ เวกเตอร์เฉพาะ เราได้นำเสนองานวิจัยขั้นสูงที่ผ่านมาที่เกี่ยวข้องกับ การสร้างคืนความละเอียดสูงยวดยิ่ง ในระดับปริภูมิ เพื่อใช้ในการรู้จำภาพวัตถุ จากรายละเอียดที่เราได้วิจารณ์งานวิจัยในเชิงวิชาการทั้งหมด เราได้นำเสนอองค์ความรู้ และอัลกอริธึมส์ใหม่ เพื่อใช้ในการรู้จำภาพวัตถุหลายอัลกอริธึมส์ ดังนี้

1. แบบจำลองทางคณิตศาสตร์สำหรับการสร้างคืนความละเอียดสูงยวดยิ่งในระดับปริภูมิย่อยสองมิติ
2. ศึกษาความเป็นไปได้ทางชีวภาพทางการมองเห็น ของ 2DPCA 2DPCA ระบุทิศทาง 2DPCA ไขว้ การสร้างคณะในการตัดสินใจ เป็นต้น
3. ระเบียบวิธีในการสร้างเมทริกซ์ความแปรปรวนร่วมเกี่ยวไขว้ อย่างเป็นระบบจากเมทริกซ์ B ซึ่งเป็นผลมาจากการแปลงเมทริกซ์ภาพวัตถุ
4. ระเบียบวิธีในการรู้จำด้วย ระบบตัวจำแนกแบบหลายตัว (MCS) ที่ทำงานร่วมกับเซตย่อยของตัวบ่งต่าง ที่สร้างมาจาก 2DPCA แบบต่างๆ

a. แบบปริภูมิย่อยสุ่ม
b. แบบมีการคัดเลือกตัวบ่งต่าง หรือ ชุดของตัวบ่งต่าง

5. องค์ความรู้ผลการแปลงเวฟเลตแบบระบุทิศ
6. การสร้างตัวบ่งต่างระบุทิศ จากผลการแปลงเวฟเลตระบุทิศ
7. ระบบตัวจำแนกแบบหลายตัว ที่ใช้กับตัวบ่งต่างระบุทิศ ที่สร้างจากผลการแปลงเวฟเลตระบุทิศ
8. ทดสอบระเบียบวิธีที่นำเสนอกับ ฐานข้อมูลภาพวัตถุมาตรฐาน เช่น FERET Yale AR ORL และ MSTAR

งานวิจัยนี้ ได้ทำการวิจัยอย่างกว้างขวางและเข้มงวด ในทุกภาคส่วนที่เกี่ยวข้องกับการวิเคราะห์ส่วนประกอบมุขสำคัญสองมิติและส่วนขยาย การสร้างคืนความละเอียดสูงยวดยิ่งในระดับปริภูมิย่อย ผลการแปลงเวฟเลตแบบระบุทิศทาง และ ความเป็นไปได้ทางชีวภาพแห่งการมองเห็น องค์ความรู้ที่ได้จะเป็นประโยชน์ต่อการวิจัย การพัฒนาการเรียน การสอน และการประยุกต์ใช้งานในการตรวจตราเพื่อรักษาความปลอดภัย

6.2 ข้อเสนอแนะ

งานวิจัยทางด้านการรู้จำใบหน้าและเป้าหมายอัตโนมัติในปัจจุบัน ยังไม่ไปถึงศักยภาพที่สามารถไปถึงได้ ในระหว่างที่ดำเนินการวิจัยนี้นั้น เราได้พบหัวข้อที่ควรถกถึงมากมาย รวมถึงการทดลองอีกเป็นจำนวนมากที่ควรจะทำ เพื่อให้เกิดความรู้ความเข้าใจที่เกี่ยวกับ ความละเอียดสูงยวดยิ่งบนฐานของเวกเตอร์เฉพาะ และ การรู้จำภาพวัตถุ ดังนี้

1. ตัวอย่างหนึ่งที่เห็นได้ชัด คือ การเราควรจะทำการทดลองความแม่นยำระหว่างระเบียบวิธีที่นำเสนอ กับ ระเบียบวิธีพื้นฐานที่ทำการรู้จำวัตถุ ซึ่งเกิดจากการรวมการตัดสินใจจากชุดของตัวจำแนกที่แต่ละตัวตัดสินใจ จากภาพความละเอียดต่ำที่แตกต่างกัน เช่นนี้ เราจะทราบสมรรถนะของ MCS เมื่อใช้ชุดของภาพ LR
2. สมรรถนะของการรู้จำ เมื่อ ระบบรู้จำที่ถูกฝึกฝนด้วย ตัวบ่งต่างที่สกัดได้จาก 2DPCA ของภาพฝึกฝนรายละเอียดสูงยวดยิ่ง เปรียบเทียบกับ ระบบรู้จำที่ถูกฝึกฝนด้วย ตัวบ่งต่างที่สกัดได้จาก การสร้างคืนความละเอียดสูงยวดยิ่งของ 2DPCA ตามสมการที่ (28) เหตุผลที่น่าจะอยู่เบื้องหลัง การใช้ตัวบ่งต่างที่ต่างกันในการฝึกฝนนั้น เป็นเพราะเราจะมีตัวบ่งต่างทั้งสองแบบในตอนต้นของการฝึกฝน ตัวบ่งต่างจะมีผลต่อนัยทั่วไปของระบบรู้จำ เมื่อทดสอบกับตัวอย่างที่ไม่เคยเห็นมาก่อน โดยเป็นการแลกเปลี่ยนกัน ระหว่างความถูกต้องในการแทนรูป กับ การสร้างคืนที่มีความรวดเร็วและมีความคงทนต่อสัญญาณรบกวน
3. การรู้จำแบบคลาสเฉพาะ ที่ทำการทดลอง อาจใช้ 2DPCA ไขว้ ร่วมเพื่อสร้างระบบรู้จำแบบหลายตัวได้ โดยการฉายภาพทดสอบลงบน เวกเตอร์เฉพาะของ 2DPCA ไขว้หลายแบบ แล้วจึงหาระยะห่างรวม
4. ในงานวิจัยที่เรานำเสนอ เป็นขบวนการสองขั้นตอน (two-stage approach) โดยในขั้นตอนแรก เราจะทำการลดขนาดของมิติ หรือ ทำการบีบอัด จากนั้นเราจะนำปริภูมิย่อยที่ได้ มาทำการสร้างคืนความละเอียดสูงยวดยิ่ง จากนั้นนำไปสร้างเซตย่อยของตัวบ่งต่าง ที่มีความซ้ำซ้อนกัน

   ในงานวิจัยทางทฤษฎีข่าวสาร มีแนวความคิดที่คล้ายคลึง แต่ทันสมัยกว่า คือ *การเข้ารหัสแหล่งกำเนิด-ช่องสัญญาณร่วม* (joint source-channel coding) ที่ทำการบีบอัดสัญญาณ และ เข้ารหัสช่องสัญญาณไปพร้อมกัน แนวความคิดทางทฤษฎีข่าวสารนี้ เป็นแนวความคิดที่ประสบความสำเร็จต่อการส่งสัญญาณภาพผ่านเครือข่ายอินเทอร์เน็ต โดยอาศัยโครงสร้างทางความคิดที่คล้ายกัน เป็นเรื่องที่น่าสนใจและท้าทายที่เราจะศึกษาค้นคว้าในหัวข้อ *การลดขนาดมิติ-ปรับ ปรุงคุณภาพร่วม* (joint dimensionality reduction-resolution enhancement) ซึ่งทำการคำนวณ

ไปพร้อมกัน แทนที่จะแยกออกเป็นสองขั้นตอน ระเบียบวิธีนี้น่าจะมีความเป็นไปได้ทางชีวภาพค่อนข้างสูง

5. การสร้างคืนปริภูมิย่อยที่นำเสนอในโครงงานวิจัยนี้ ได้ตั้งสมมติฐานให้การกระจายตัวของปริภูมิย่อยของใบหน้า ในแต่ละปริภูมิย่อยมีการกระจายตัวแบบเกาส์ ซึ่งอาจเป็นการกระจายตัวที่ไม่ถูกต้องเท่าใดนัก ผู้วิจัยคิดว่าในเวกเตอร์เฉพาะที่มีขนาดใหญ่ไม่มาก แต่มีความสำคัญในการวิจัย น่าจะมีการกระจายตัวแบบอื่น เช่น การกระจายตัวแบบลาปราส หรือ เพียร์สัน เป็นต้น

   หากดำเนินการทดลองแล้วพบว่า การกระจายตัวของค่าบ่งต่างของปริภูมิย่อยมีการกระจายตัวที่ต่างออกไปแล้ว เราจะสามารถคำนวณการประมาณ MAP (maximum a posteriori estimation) เพื่อสร้างสมการใหม่ในการสร้างคืนปริภูมิความละเอียดสูงยวดยิ่ง ที่ยังไม่มีใครคิดมาก่อน

6. PCA และ 2DPCA ให้ตัวบ่งต่างที่มี อำนาจในการจำแนก (discriminant power) ระหว่างคลาสต่ำ หากมีการนำ LDA และ 2DLDA มาร่วมในงานวิจัยจะทำให้ความแม่นยำของการรู้จำดีขึ้น ยกตัวอย่างเช่น LDA ของ PCA ซึ่งเป็นแนว ความคิดที่ง่ายดายที่ยังไม่มีนักวิจัยท่านใดคิดการสร้างคืนความละเอียดสูงยวดยิ่งในระดับปริภูมิ

7. 2DLDA ของ 2DPCA ได้มีการนำเสนอโดยคณะผู้วิจัย (Sanguansat, 2006) แต่ยังไม่มีการสร้างคืนความละเอียดสูงยวดยิ่งในระดับปริภูมิ โดยการแก้ไขงานวิจัยที่มีอยู่เพียงเล็กน้อยเราจะสามารถเพิ่ม อำนาจในการจำแนก ให้กับระบบรู้จำที่เราสร้างขึ้นได้

8. การค้นหาการแปลงเมทริกซ์ B แบบอื่นๆ ที่มีความเหมาะสม และ มีความเป็นไปได้ทางชีวภาพ เพื่อใช้ในการสร้างเซตย่อยสำหรับกำเนิด ตัวบ่งต่างจำนวนมาก ความบ่งต่างที่ได้จากการคำนวณความแปรปรวนร่วมเกี่ยวไขว้ ควรจะมีสหสัมพันธ์ ระหว่างคู่พิกเซล ที่ให้อำนาจในการจำแนกสูง หรือ มีความหลากหลาย โดยอาจสร้างการแปลงเมทริกซ์ B หลายชนิด บางชนิดเน้นไปที่ สหสัมพันธ์ที่เน้นขอบภาพ บางชนิดเน้นไปที่ความสัมพันธ์ระหว่างอวัยวะ เช่น ตา กับ ปาก เป็นต้น

9. บทความเกี่ยวกับ ระบบตัวจำแนกแบบหลายตัว กล่าวถึง การรวมการตัดสินใจ ซึ่งเป็นหนึ่งในสองส่วนที่สำคัญของระบบ ว่า การลงคะแนนเสียง เสมือนเป็นการกรองผ่านต่ำสัญญาณรบกวนออกจากระบบ ผลการแปลงเวฟเลตมีผลตอบ สนองทางความถี่ทั้งในส่วนของการกรองผ่านต่ำ และ การกรองผ่านแถบความถี่ ดังนั้นการนำแนวความคิด ผลการแปลงเวฟเลตหนึ่งมิติ มาร่วมใช้ในการหา *กฎการรวม* (sum rule) ของการตัดสินใจ เพื่อให้ได้ ค่าถ่วงน้ำหนักจำเพาะ ที่เหมาะสม ดังต่อไปนี้

a. แถบความถี่ย่อยด้านกรองต่ำ
b. การขจัดสัญญาณรบกวน (denoising)
c. การใช้ฟังก์ชันฐานหลักอำนาจจำแนกท้องถิ่น (local discriminant basis)
d. *การวิเคราะห์รูปแบบแบบหลายการพรรณนา* (multiple description pattern analysis) (Asdornwised, 2005) ซึ่งเป็น ส่วนขยายของ *การเข้ารหัสแบบหลายการพรรณา* (multiple description coding) ซึ่งเป็นที่นิยมมากในการบีบอัดสัญญาณภาพเพื่อส่งในข่ายอินเทอร์เน็ต

เมื่อเราสามารถเลือกแถบความถี่ย่อยที่เหมาะสมได้แล้ว เราจะทำการสร้างคืน ผลการแปลงที่ใช้ให้กับมาที่มิติเดิม ซึ่งเราจะได้กฎการรวม ที่ดีที่สุด

10. เร็วๆนี้ ได้มีงานวิจัยในการประยุกต์ใช้ *การสุ่มสัญญาณที่ถูกบีบอัด* (compressed sensing) ในการรู้จำใบหน้า แนวความคิดที่กล่าวมาข้างต้นทั้งหมดสามารถนำมาพัฒนาไปในแนวทางนี้ได้

จะเห็นได้ว่างานวิจัยนี้ได้ก่อให้เกิดหัวข้อวิจัยเปิดขึ้นอีกมากมาย งานวิจัยที่ดำเนินการต่อเนื่องจากนี้ไปจะก่อให้เกิดผลกระทบในเชิงวิชาการในวงกว้างต่อไปในอนาคต

ภาคผนวก ก.

ผลงานที่ได้รับ

I. Book Chapters

1. **Widhyakorn Asdornwised**, “Face and Automatic Target Recognition based on Super-Resolved Discriminant Subspace,” in Peter M. Corcoran (Ed.), *Reviews, Refinements and New Ideas Face Recognition*, pp. 167-180, InTech Open Access Publisher, ISBN 978-953-307-368-2, Hard cover, 328 pages, 2011.

II. International periodical journals

--

III. International Conference Proceedings

1. **Widhyakorn Asdornwised,** “Face and automatic target recognition based on class-specific superresolution subspace,” Defense and Security Symposium 2007, Proc. of SPIE, Vol. 6567, Orlando, Florida, 2007.
2. Parinya Sa-ngungsat, **Widhyakorn Asdornwised**, Sanparith Marukatat, and Somchai Jitapunkul, “Image Cross-Covariance Analysis for Face Recognition,” IEEE TENCON, Taipei, 30 Oct. -2 Nov. 2007.
3. Parinya Sa-ngungsat, **Widhyakorn Asdornwised**, Sanparith Marukatat, and Somchai Jitapunkul, “Two-Dimensional Random Subspace Analysis for Face Recognition,” ISCIT 2007, 2007 International Symposium Communications and Information Technology, Sydney, 16-19 Oct. 2007.
4. Parinya Sa-ngungsat, **Widhyakorn Asdornwised**, Sanparith Marukatat, and Somchai Jitapunkul, “Two-Dimensional Diagonal Random Subspace Analysis for Face Recognition,” 2007 International Conference on Telecommunications, Industry, and Regulatory Development, Bangkok, pp. 66- 69, 19-21 Aug. 2007.

ภาคผนวก ข.

ชุดคำสั่งเทียม

% function cross-covariance matrix

*S*1: For i = 1 to the number of classifiers

*S*2: Transforming all training images A to B

*S*3: Find the outer product $B^T A$

*S*4: A=B

*S*5: Construct the i^{th} 2DPCA

*S*6: End

% function super-resolved 2DPCA

*S*1: Transforming all LR training images A_{LR} to diagonal images LR $A_{d,\,LR}$

*S*2: Find the LR outer product $A_{d,LR}^T A_{d,LR}$

*S*3: Find LR 2DPCA eigenvectors

*S*4: Transforming all HR training images A_{HR} to diagonal images HR $A_{d,HR}$

*S*5: Find the HR outer product $A_{d,HR}^T A_{d,HR}$

*S*6: Find HR 2DPCA eigenvectors

*S*7: Reconstruct super-resolved feature matrix using (28)

% function Diagonal-2DPCA with Random Subspace Method

*S*1: Transforming all training images A to diagonal images A_d

*S*2: Find the outer product $A_d^T A_d$

*S*3: Project image, A onto the calculated 2DPCA

*S*4: For i = 1 to the number of classifiers

*S*5: Randomly select a r dimensional random subspace from Y ($r < m$).

*S*6: Construct the i^{th} nearest neighbor classifier using the proposed distance

*S*7: End

*S*8: Combine all of the outputs of classifiers by using majority voting

% function Directional-2DPCA with Feature Selection Method

*S*1: For i = 1 to the number of classifiers

*S*2: Construct the i^{th} 2DPCA from cross-covariance matrix function

*S*3: Construct the i^{th} nearest neighbor classifier using the proposed distance

*S*4: End

*S*5: Select K using Kullback-Leibler distance or Mutual Information cost function

S6: Combine all of the outputs of classifiers by using majority voting or sum rule

% function 2DPCA with Real Dual-Tree Wavelet

*S*1: Implement dualtree2D for each training image

S1: For i = 1 to the number of classifiers

*S*2: Construct the i^{th} 2DPCA from the i^{th} directional subband

*S*3: Construct the i^{th} nearest neighbor classifier using the proposed distance

*S*4: End

*S*5: Select K using Kullback-Leibler distance or Mutual Information cost function

S6: Combine all of the outputs of classifiers by using majority voting or sum rule

เอกสารอ้างอิง

Park, S. C.; Park, M. K. & Kang, M. G. (2003). Super-Resolution Image Reconstruction: A Technical Overview. *IEEE Signal Processing Magazine*, Vol. 20, pp. 21-36

Belhumeur, P. N.; Hespanha, J. P. & Kriegman, D. J. (1997). Eigenface vs. Fisherfaces: Recognition using Class Specific Linear Projection, *IEEE Trans. Pattern Anal. Machine Intell*, Vol.19 (May 1997), pp. 711-720

Nguyen, N.; Milanfar, P. & Golub, G. H (2001). A Computationally Efficient Image Super-Resolution Algorithm. *IEEE Trans. Image Proc.*, Vol. 4, pp. 573-583

Lin, F.; Cook, J.; Chandran, V. & Sridharan, S. (2005). Face Recognition from Super-Resolved Images, *ISSPA*, Australia, August 2005

Wagner, R.; Waagen, D. & Cassabaum, M. (2004). Image Super-Resolution for Improved Automatic Target Recognition, *SPIE Defense & Security Symposium*, Orlando, USA, April 2004

Gunturk, B. K.; Batur, A. U.; Altunbasak, Y.; Hayes, M. H. & Mersereau, R. M. (2003). Eigenface-Domain Super-Resolution for Face Recognition. *IEEE Trans. Pattern Anal. and Mach. Intell.*, Vol. 12, pp. 597-606

Jia, K. & Gong, S. (2005). Multi-Modal Tensor Face for Simultaneous Super-Resolution and Recognition, *Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV)*, pp. 1683-1690, Washington, DC, USA, 2005

Sezer, O.G.; Altunbasak, Y. & Ercil, A. (2006). Face Recognition with Independent Component-based Super-Resolution. *Proc. of SPIE: Visual Communications and Image Processing*, Vol. 6077, 2006

Werblin, F. & Roska, B. (2007) The Movies in Our Eyes, Scientific American, Vol. 296, No. 4, pp. 72-79.

Fukunnaga, K. (1991). *Introduction to Statistical Pattern Recognition*, 2nd ed. New York, NY: Academic.

Turk, M. & Pentland, A. (1991). “Eigenfaces for Recognition,” *J. of Cognitive Neuro-science*, Vol. 3, no. 1, pp. 71–86.

PCA (n.d.). Principal Component Analysis. Available from http://en.wikipedia.org/wiki/Principal_component_analysis

Shan, S.; Gao, W. & Zhao, D. (2003). Face Recognition based on Face-Specific subspace. *International Journal of Imaging Systems and Technology*, Vol.13, No.1, pp.23-32

Zhao, W.; Chellappa, R. & Krishnaswamy, A. (1998) Discriminant Analysis of Principal Components for Face Recognition, *Proceedings of the 3rd IEEE International Conference on Face and Gesture Recognition (FG)*, pp. 336-341, Nara, Japan, 14-16 April 1998

Yang, J.; Zhang, D.; Frangi, A.F. & Yang, J. yu (2004) Two-dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition. *IEEE Trans. Pattern Anal. and Mach. Intell.*, Vol.26, pp.131-137, Jan. 2004.

Occam (n.d.). Occam's Razor. Available from http://en.wikipedia.org/wiki/Occam's_razor

Bellman, R. (1961). Adaptive Control Processes. Princeton University Press, Princeton, N.J.

Zhang, D. & Zhou, Z.-H. (n.d.) $(2D)^2$PCA: 2-directional 2-dimensional PCA for Efficient Face Representation and Recognition. *Neurocomputing*.

Shepard, R.N. & Hurwitz, S. (1985). Upward Direction, Mental Rotation, and Discrimination of lLeft and Right Turns in Maps. in *Visual cognition*. Steven Pinker Ed., pp. 161-192, MIT Press, Cambridge, Mass., 1985.

Edelman, S. (1992). A Neural Model of Object Recognition in Human Vision. in *Neural Networks for Perception.* Harry Wechsler Ed., pp. 25-40, Academic Press, Boston, 1992.

Polikar, R. (2006). Ensemble Based System in Decision Making. *IEEE Circuits and Systems Magazine,* Third Quarter 2006, pp. 21-45.

Wolf, L. & Bileschi, S. (2005). Combining variable selection with dimensional reduction. AI memo 2005-009, CBCL Memo 247, March 2005.

Guyon, I. & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research. Vol. 3, pp. 1157-1182, 2003.

Saito, N. & Coifman, R. R. (1995). Local Discriminant Bases and Their Applications. *Journal of Mathematical Imaging and Vision.* No. 5, Kluwer Academic Publishers, pp. 337-358.

Bamberger R.H. & Smith, M.J.T. (1992). A Filter Bank for the Directional Decomposition of Images: Theory and Design. *IEEE Trans. Signal Proc.,* Vol. 40, No. 4, April 1992, pp.882-893.

Chung, Ki-Chung; Kee, S.C. & Kim, S. R. (1999). Face recognition using principal component analysis of Gabor filter responses. Int. Workshop on Recog. Anal. And Tracking of Face and Gestures in Real-Time Systems, pp. 53-57, 1999.

Nguyen, T. T. & Oraintara, S. (2007).A class of multiresolution filter banks. *IEEE Trans. on Signal Proc.*, Vol. 55, No. 3, pp. 949 - 961, March 2007.

Durand, S. (2008). On the Construction of Discrete Directional Wavelets: HDWT, Space-Frequency Localization, and Redundancy Factor. Multiscale Modeling & Simulation 7(3), pp. 1325-1347, 2008.

Selesnick (n.d.). Wavelet Software at Brooklyn Polytechnics. Available from http://taco.poly.edu/WaveletSoftware/dt2D.html

Asdornwised, Widhyakorn & Jitapunkul, Somchai. (2002). Multiresolution-Based Committees of Networks: A Bayesian Point of View. IEEE ICIT '02. 2002 IEEE International Conference on Industrial Technology. Vol. 1, pp. 643-648, 2002.

Sermwuthisarn, Parichat; Asdornwised, Widhyakorn; Jitapunkul, Somchai & Marukatat, S. (2007). Decision Combination of Multiple Classifier Systems for Multiresolution Face Recognition. ECTI-CON 2007, Chiang Rai, Thailand, 9-12 May 2007.

Sermwuthisarn, P.; Asdornwised, W.; Jitapunkul, S. & Marukatat, S. (2008). Information-Theoretic Wavelet Subband Selection for Face Recognition. ECTI-CON 2008, Krabi, Thailand, 14-17 May 2008.

Kong, H.; Li, X.; Wang, L.; Teoh, E. K.; Wang, J.-G. & Venkateswarlu, R. (2005). Generalized 2D Principal Component Analysis," *IEEE International Joint Conference on Neural Networks (IJCNN)*.

Sanguansat, P.; Asdornwised, W.; Jitapunkul, S. & Marukatat, S. (2006). Two-Dimensional Linear Discriminant Analysis of Principle Component Vectors for Face Recognition. *IEICE Transaction on Information and System: Special Issue on Machine Vision Applications,* Vol. E89- D, No. 7, pp. 2164-2170

Sanguansat, P.; Asdornwised, W.; Jitapunkul, S. & Marukatat, S. (2007a). Image Cross-Covariance Analysis for Face Recognition. IEEE TENCON, Taipei, 30 Oct.-2 Nov. 2007.

Asdornwised, W. (2008) Personal Communications.

Kumar, V. B. G. & Aravind, R. (2008). A 2D Model for Face Superresolution. *Proceedings of the 17th International Conference on Pattern Recognition (ICPR)*, pp. 1–4, Tampa, Florida, USA, 2008

Gsmooth (n.d.). Gaussian Smoothing. Available from http://homepages.inf.ed.ac.uk/rbf/HIPR2/gsmooth.htm

Schott, J. R. (2005), *Matrix Analysis for Statistics,* John Wiley & Sons, ISBN 0-471-66983-0, New Jersey, USA

Ho, T.K. (1998). Nearest neighbors in random subspaces. in *Proceedings of the 2nd Int'l Workshop on Statistical Techniques in Pattern Recognition*, Sydney, Australia, August 1998, pp. 640–648.

Chawla, N.V. & Bowyer, K. (2005). Random subspaces and subsampling for 2-d face recognition. *Computer Vision and Pattern Recognition*, vol. 2, 2005, pp. 582–589.

Wang, X. & Tang, X. (2006). Random Sampling for Subspace Face Recognition. International Journal of Computer Vision. Vol. 70. No. 1, pp. 91–104, 2006.

Nguyen, N.; Liu, W. & Venkatesh, S. (2007). Random Subspace Two-Dimensional PCA for Face Recognition. In Proceedings of PCM'2007. pp.655~664

Sanguansat, P.; Asdornwised, W.; Marukatat, S. & Jitapunkul, S. (2007b). Two-Dimensional Random Subspace Analysis for Face Recognition. In International Symposium on Communications and Information Technologies (ISCIT), Sydney, Australia, Oct. 2007.

Sanguansat, P.; Asdornwised, W.; Marukatat, S. & Jitapunkul, S. (2007c). Two-Dimensional Diagonal Random Subspace Analysis for Face Recognition. In International Conference on Telecommunications, Industry and Regulatory Development, Bangkok, Thailand, pp. 66-69, August 2007.

Tabassian, M.; Ghaderi, R. & Ebrahimpour, R. (2010). Subspace Ensembles for Face Recognition. Australian Journal of Intelligent Information Processing Systems, Vol. 12, No. 3, 2010.

Phillips, P. J.; Moon, H.; Rauss, P. J.; & Rizvi, S. (2000). The FERET evaluation methodology for face recognition algorithms. *IEEE Trans. on Pattern Analysis and Machine Intelligence*. Vol. 22, No. 10, October 2000.

Yale University (1997). The Yale Face Database. Available from http://cvc.yale.edu/projects/yalefaces/yalefaces.html.

Martinez, A. & Benavente, R. (1998). The AR Face Database. Available from http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html.

The Center for Imaging Science (CIS) (1997). DARPA Moving and Stationary Target Recognition Program (MSTAR). Available from http://cis.jhu.edu/data.sets/MSTAR.

Asdornwised, W. & Jitapunkul, S. (2005). Multiple Description Pattern Analysis: Robustness to Misclassification using Local Discriminant Frame Expansions. *IEICE Trans. Inf. & Syst.*, Vol. E88-D, No. 10, pp. 2296-2307, Oct. 2005.

Chapter Number

Face and Automatic Target Recognition Based on Super-Resolved Discriminant Subspace

Widhyakorn Asdornwised
Department of Electrical Engineering, Chulalongkorn University, Bangkok Thailand

1. Introduction

Recently, super-resolution reconstruction (SRR) method of low-dimensional face subspaces has been proposed for face recognition. This face subspace, also known as *eigenface,* is extracted using *principal component analysis* (PCA). One of the disadvantages of the reconstructed features obtained from the super-resolution face subspace is that no class information is included. To remedy the mentioned problem, at first, this chapter will be discussed about two novel methods for super-resolution reconstruction of discriminative features, i.e., *class-specific* and *discriminant analysis of principal components*; that aims on improving the discriminant power of the recognition systems. Next, we discuss about *two-dimensional principal component analysis* (2DPCA), also refered to as *image PCA*. We suggest new reconstruction algorithm based on the replacement of PCA with 2DPCA in extracting super-resolution subspace for face and automatic target recognition. Our experimental results on Yale and ORL face databases are very encouraging. Furthermore, the performance of our proposed approach on the MSTAR database is also tested.

In general, the fidelity of data, feature extraction, discriminant analysis, and classification rule are four basic elements in face and target recognition systems. One of the efficacies of recognition systems could be improved by enhancing the fidelity of the noisy, blurred, and undersampled images that are captured by the surveillance imagers. Regarding to the fidelity of data, when the resolution of the captured image is too small, the quality of the detail information becomes too limited, leading to severely poor decisions in most of the existing recognition systems. Having used super-resolution reconstruction algorithms (Park et al., 2003), it is fortunately to learn that a high-resolution (HR) image can be reconstructed from an undersampled image sequence obtained from the original scene with pixel displacements among images. This HR image is then used to input to the recognition system in order to improve the recognition performance. In fact, super-resolution can be considered as the numerical and regularization study of the ill-conditioned large scale problem given to describe the relationship between low-resolution (LR) and HR pixels (Nguyen et al., 2001).

On the one hand, feature extraction aims at reducing the dimensionality of face or target image so that the extracted feature is as representative as possible. On the other hand, super-resolution aims at visually increasing the dimensionality of face or target image. Having applied super-resolution methods at pixel domain (Lin et al., 2005; Wagner et al., 2004), the performance of face and target recognition applicably increases. However, with the emphases on improving computational complexity and robustness to registration error

and noise, the continuing research direction of face recognition is now focusing on using eigenface super-resolution (Gunturk et al., 2003; Jia & Gong, 2005; Sezer et al., 2006).
The essential idea of eigen-domain based super-resolution using 2D eigenface instead of the conventional 1D eigenface is to overcome the three major problems in face recognition system, i.e., the curse of dimensionality, the prohibited computing processing of the singular value decomposition at visually improved high-quality image, and natural structure and correlation breaking in the original data.
In Section 2, the basic of super-resolution for low-dimensional framework is briefly explained. Then, discriminant approaches are detailed in Section 3 with the purpose of increasing the discrimination power of the eigen-domain based super-resolution. In Section 4, the implement of the two dimensional eigen-domain based super-resolution is addressed. We also discuss the possibility of the extension of two dimensional eigen-domain based super-resolution with discriminant information in Section 5. Finally, Section 6 provides the experimental results on the Yale and ORL face databases and MSTAR non-face database.

2. Eigenface-domain super-resolution

The fundamental of the super-resolution for in low-dimensional face subspace is formulated here. The important of the image super-resolution model and its eigenface-domain based reconstruction is that they can be used for practical extensions of one- and two-dimensional super-resolved discriminant face subspaces in the next sections, respectively.

2.1 Image super-resolution model

According to the numerically computational SRR framework (Nguyen et al., 2001), the relationship between an HR image and a set of LR images can be formulated in matrix form as follows:

$$\mathbf{f}_k = D_k B_k E_k \mathbf{x} + \mathbf{n}_k, \quad 1 \le k \le p \tag{1}$$

where $\boldsymbol{p}$ is the number of available frame, $\mathbf{f}_k$ and $\mathbf{x}$ are vectors extracted from the k^{th} LR image frame and HR image in lexicographical order, respectively, and D is the down-sampling operator, B is the blurring or averaging operator, and E_k is the affine transform, and $\mathbf{n}_k$ is noise of the frame k, respectively.
Thus, we can reformulate (1) as

$$\begin{bmatrix} \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_p \end{bmatrix} = \begin{bmatrix} D_1 B_1 E_1 \\ \vdots \\ D_p B_p E_p \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{n}_1 \\ \vdots \\ \mathbf{n}_p \end{bmatrix} \tag{2}$$

or

$$\begin{aligned} \mathbf{f}_k &= H_k \mathbf{x} + \mathbf{n}_k \\ \mathbf{f} &= \mathbf{H}\mathbf{x} + \mathbf{n}. \end{aligned} \tag{3}$$

The above equation can be solved as an inverse problem with a regularization term, or

$$\mathbf{x} = \mathbf{H}^{\mathrm{T}} (\mathbf{H}\mathbf{H}^{\mathrm{T}} + \lambda \mathbf{I})^{-1} \mathbf{f}. \tag{4}$$

It should be noted that the matrix **H** is a very large sparse matrix. As a result, the analytic solution of **x** is very hard to find. One of the popular methods used for finding the solution of this kind of the inverse problem is by using conjugate gradient method.

2.1 Reconstruction algorithm

Common preprocessing step used for pattern recognition and in compression schemes is dimensionality reduction of data. In image analysis, PCA is one of the popular methods used for dimensionality reduction. Let $\mathbf{\Phi}$ be an optimal eigenface that removes the redundancy by decorrelating the image data **x**. The optimal eigenfaces are coded in its columns. Face image **x** is assumed to be vectored. Thus, the optimal image representation of **x** can be written as

$$\mathbf{x} = \mathbf{\Phi a} + \mathbf{e}_x , \tag{5}$$

where **a** is the $L \times 1$ dimensional feature that represents **x**, and $\mathbf{e_x}$ is its representation error. Given that $\mathbf{\Psi}$ is the $\beta^2 N^2 \times L$ matrix that contains eigenfaces of the k^{th} LR image frame, where the scaling resolution factor β is within the range 0 to 1 and N is the total face image pixels. We can formulate the low-resolution image representation as

$$\mathbf{f}_k = \mathbf{\Psi}\hat{\mathbf{a}}_k + \mathbf{e}_{f_k} . \tag{6}$$

By substituting (5) and (6) in (3), we obtain

$$\mathbf{\Psi}\hat{\mathbf{a}}_k + \mathbf{e}_{f_k} = H_k \mathbf{\Phi a} + H_k \mathbf{e}_x + \mathbf{n}_k . \tag{7}$$

Since

$$\mathbf{\Psi}^T \mathbf{e}_{f_k} = 0 \tag{8}$$

and

$$\mathbf{\Psi}^T \mathbf{\Psi} = 1 . \tag{9}$$

It is easy to derive the following equation

$$\hat{\mathbf{a}}_k = \mathbf{\Psi}^T H_k \mathbf{\Phi a} + \mathbf{\Psi}^T H_k \mathbf{e}_x + \mathbf{\Psi}^T \mathbf{n}_k . \tag{10}$$

By considering the second and third terms as the observation noise with Gaussian distribution (Gunturk et al., 2003), we can obtain

$$\hat{\mathbf{a}}_k = \Lambda_k \mathbf{a} + \mathbf{\Psi}^T \zeta_k , \tag{11}$$

where

$$\zeta_k = H_k \mathbf{e}_x + \mathbf{n}_k , \tag{12}$$

and

$$\Lambda_k = \mathbf{\Psi}^T H_k \mathbf{\Phi} . \tag{13}$$

Without loss of generality, we can numerically solve for the true super-resolution feature vector at the eigen-domain level as in (5), or

$$\mathbf{a} = \mathbf{\Lambda}^{\mathrm{T}}(\mathbf{\Lambda}\mathbf{\Lambda}^{\mathrm{T}} + \gamma\mathbf{I})^{-1}\hat{\mathbf{a}}\,, \tag{14}$$

where γ is the regularization term. In particular, we introduce the notation p in (14) in order to differentiate the PCA-domain based super-resolution approach (Gunturk et al., 2003) from our proposed approaches which will be presented in the upcoming sections,

$$\mathbf{a}_p = \mathbf{\Lambda}_p^T(\mathbf{\Lambda}_p\mathbf{\Lambda}_p^T + \gamma\mathbf{I})^{-1}\hat{\mathbf{a}}_p\,. \tag{15}$$

3. Discriminant face subspaces

PCA and its eigenface extension are constructed around the criteria of preserving the data distribution. Hence, it is well suited for face representation and reconstruction from the projected face feature. However, it is not an efficient classification method because the between classes relationship has been neglected. Here, we discuss on the possibilities that how we can embed discriminant information into eigenface-domain based super-resolution.

3.1 Face-specific subspace Super-resolution

As widely known, the eigen-domain based face recognition methods use the subspace projections that do not consider class label information. The eigenface's criterion chooses the face subspace (coordinates) as the function of data distribution that yields the maximum covariance of all sample data. In fact, the coordinates that maximize the scatter of the data from all training samples might not be so adequate to discriminate classes. In recognition task, a projection is always preferred to include discrimination information between classes. One of the extensions of eigenface, called face-specific subspace (FSS) (Shan, 2003), is proposed as an alternative feature extraction method to include class information for face recognition application. According to FSS, each reduced dimensional basis of class-specific subspace (CSS) is learned from the training samples of the same class. Actually, each individual set of CSS optimally represents the data within its own class with negligible error. As a result, large representation error occurs, when the input data is projected and then reconstructed using a reduced set with less maximum covariance coordinates (or equivalently, using a set of principal components that does not belong to the input class). This way, by using reconstruction error obtained from projection-reconstruction process between classes, also called distance from CSS (DFCSS), a new metric can be suitably used as the distance for classifying the input data. In other words, the smaller the DFCSS is, the higher the probability that the input data belongs to the corresponding class will be. Similar work based on FSS (Belhumeur, 1997) attacking wide attentions in face recognition society is also published recently.

The original face-specific subspace (FSS) was proposed to manipulate the conventional eigenface in order to improve the recognition performance. According to FSS, the difference between FSS and the traditional method is that the covariance matrix of the p^{th} class is individually evaluated from training samples of the p^{th} class. Thus, the p^{th} FSS is represented as a 4-tuple, i.e., the projection matrix, the mean of the p^{th} class, the eigenvalues of covariance matrix, and the dimension of the p^{th} CSS. For identification, the input sample is

projected using all CSSs and then reconstruct by those CSSs. If reconstruction error which obtained from the p^{th} CSS is minimum then the input sample is belong to the p^{th} class, also called distance from CSS (DFCSS).

There are many advantages of using CSS in face and target recognition. For example, the transformation matrices are trained from samples within their own classes, thus it is more optimum (using fewer components) to represent each sample in its own class than a transformation matrix trained by samples in all classes. Additionally, since DFCSS is the distance between the original image and its reconstruction image obtained from CSS, the memory space needed is only for storing the C transformation matrices, where C is the number of classes. This is far less than the conventional subspace methods, where we need to store both a single all-classes transformation matrix and also its prototypes (a large set of feature vectors calculated for all training samples). Moreover, the number of distance calculation in CSS is less than the number of distance calculation in conventional methods, since the number of classes is usually less than the number of training samples.

By combining super-resolution reconstruction approach with class-specific idea, a new method for face and automatic target recognition is proposed.

3.2 Discriminant analysis of principal components

The PCA's criterion chooses the subspace as the function of data probability distribution while *linear discriminant analysis* (LDA) chooses the subspace which yields maximal inter-class distance, and at the same time, keeping the intra-class distance small. In general, LDA extracts features which are better suitable for classification task. Both techniques intend to project the vector representing face image onto lower dimensional subspace, in which each 2D face image matrix must be first transformed into vector and then a collection of the transformed face vectors are concatenated into a matrix.

The PCA and LDA implementation causes three major problems in pattern recognition. First of all, the covariance matrix, which collects the feature vectors with high dimension, will lead to *curse of dimensionality*. It will further cause the very demanding computation both in terms of memory and time. Secondly, the spatial structure information could be lost when the column-stacking vectorization and image resize are applied. Finally, especially in face recognition task, the available number of training samples is relatively small compared to the feature dimension, so the covariance matrix which estimated by these features trends to be singular, which is addressed ased *singularity problem* or *small sample zize* (SSS) problem. Especially, as a supervised technique, LDA has a tendency to overfitting because of the SSS problems.

Various solutions have been proposed for solving the SSS problem. Among these LDA extensions, Fisherface and the discriminant analysis of principal components framework (Zhao, 1998) demonstrate a significant improvement when applying LDA over principal components subspace. Since both PCA and LDA can overcome the drawbacks of each other.

It has also been noted that LDA faces two certain drawbacks when directly applied to the original input space. First of all, some non-face information such as image background has been regarded by LDA as the discriminant information. This causes misclassification when the face of the same subject is presented on different background. Secondly, the within-class scatter matrix trends to be singular when SSS problem has occurred. Projecting the high dimensional input space into low dimensional subspace via PCA first can solve the shortcomings of the LDA problems. In other words, class information should be included to PCA by incorporating LDA.

3.2.1 Proposed reconstruction algorithm

Here, we can obtain a linear projection which maps the HR input image **x** first into the face subspace, and finally into the classification space **z**. Thus, we can modify the equation (5) to be

$$\mathbf{x} = \mathbf{W}_z \mathbf{\Phi} \mathbf{a}_{dp} + \mathbf{e}_x + \mathbf{e}_z , \tag{16}$$

where $\mathbf{W}_z$ is the optimal discrimination projection obtained from solving the generalized eigenvalue problem:

$$\mathbf{S}_\mathbf{B} \mathbf{W}_z = \lambda \mathbf{S}_\mathbf{w} \mathbf{W}_z , \tag{17}$$

and $\mathbf{S}_\mathbf{B}$, $\mathbf{S}_\mathbf{W}$ are the *between-class* and *within-class scatter matrices*, respectively. Similarly, we can find the optimal discriminant project of the LR image frame $\mathbf{W}_z^{'}$, by little manipulating on (16)-(17) with corresponding LR images.

With little manipulations, we can reconstruct discriminant analysis of principal components based super-resolution as

$$\mathbf{a}_{dp} = \mathbf{\Lambda}_{dp}^T (\mathbf{\Lambda}_{dp} \mathbf{\Lambda}_{dp}^T + \gamma \mathbf{I})^{-1} \hat{\mathbf{a}}_{dp} , \tag{18}$$

where

$$\mathbf{\Lambda}_{lp} = \mathbf{W}_z^{'} \mathbf{\Psi}^\mathrm{T} \mathbf{H} \mathbf{W}_z \mathbf{\Phi} , \tag{19}$$

and

$$\boldsymbol{\zeta} = \mathbf{H}\mathbf{e}_\mathbf{x} + \mathbf{H}\mathbf{e}_z + \mathbf{n}_k . \tag{20}$$

4. Two-dimensional eigen-domain based super-resolution

Recently, Yang (Yang et al., 2004) proposed an original technique called *two-dimensional principal component analysis* (2DPCA), in which the image covariance matrix is computed directly on image matrices so the spatial structure information can be preserved. One of the benefits of this method is that the dimension of the covariance matrix just equals to the width of the face image or the height in case of 2DPCA variant. This size is much smaller than the size of covariance matrix estimated in PCA. Therefore, the image covariance matrix can be better estimated with full rank in case of few training examples, like in face recognition.

We now consider linear projection of the form

$$\tilde{\mathbf{x}} = \mathbf{\Theta} \mathbf{v} + \tilde{\mathbf{e}}_\mathbf{x} , \tag{21}$$

where $\tilde{\mathbf{x}}$ represents any face image in its original matrix form, $\{\theta_1, \cdots, \theta_d\}$, be the d largest eigenvectors that can be form to be $\mathbf{\Theta}$, and $\mathbf{v}$ is the projected HR feature of this image on $\mathbf{\Theta}$, called *principal component matrix*. The criterion used for obtaining the eigenvectors in (21) has been descriptively shown in Yang and Sanguangsat (Yang et al., 2004; Sanguangsat, 2006).

4.1 Alternative image super-resolution model

LR and HR images can be simply related as (Vijay, 2008)

$$\tilde{\mathbf{f}}_k = L_k \tilde{\mathbf{x}} R_k + \tilde{\mathbf{n}}_k, \quad 1 \le k \le p. \tag{22}$$

where $\boldsymbol{p}$ is the number of available frame; L_k , R_k are downsampling matrices, and $\tilde{\mathbf{f}}_k$, $\tilde{\mathbf{x}}$ are image matrices from the k^{th} LR image frame and HR image, respectively. It should be noted that two-dimensional Gaussian blur can be represented by using together the two separate L_k and R_k . An extension to downsampling and affine transform can also be easily conducted by placing the elements of the matrices properly (Gsmooth, n.d.). It should also be noted that both the input LR and HR image are represented in its original matrix form. We do not transform the LR and HR images to be vectors in lexicography order as in (1).

4.2 Proposed reconstruction algorithm

Thus,

$$\boldsymbol{\Gamma}\hat{\mathbf{v}}_k + \mathbf{e}_{\tilde{f}_k} = L_k \boldsymbol{\Theta} \mathbf{v} R_k + L_k \mathbf{e}_x R_k + \tilde{\mathbf{n}}_k, \tag{23}$$

where $\{\Gamma_1, \cdots, \Gamma_d\}$, be the d largest eigenvectors that can be form to be $\boldsymbol{\Gamma}$,and $\hat{\mathbf{v}}_k$ is the projected LR feature of the image on $\boldsymbol{\Gamma}$.
Without loss of generality,

$$\boldsymbol{\Gamma}^T \mathbf{e}_{f_k} = 0 \tag{24}$$

and

$$\boldsymbol{\Gamma}^T \boldsymbol{\Gamma} = 1. \tag{25}$$

It is easy to derive the following equation

$$\hat{\mathbf{v}}_k = \boldsymbol{\Gamma}^T L_k \boldsymbol{\Theta} \mathbf{v} R_k + \boldsymbol{\Gamma}^T L_k \mathbf{e}_x R_k + \boldsymbol{\Gamma}^T \tilde{\mathbf{n}}_k. \tag{26}$$

It should be noted that $\hat{\mathbf{v}}_k$ is a feature matrix, unlike $\hat{\mathbf{a}}_k$ which is a feature vector. Thus, it is a little more complicated to solve the inverse problem for super-resolution feature matrix $\mathbf{v}_k$.
By applying vector operator as presented in Kumar and Schott (Kumar, 2008; Schott, 2005), (26) can be rewritten as

$$\hat{\beta}_k = \Xi_k \beta + \eta_k, \tag{27}$$

where $\hat{\beta}_k = vec(\hat{\mathbf{v}}_k)$, $\beta_k = vec(\mathbf{v})$, $\Xi_k = R_k^T \otimes \boldsymbol{\Gamma}^T L_k \boldsymbol{\Theta}$ and $\eta_k = vec(\boldsymbol{\Gamma}^T L_k \mathbf{e}_x R_k + \boldsymbol{\Gamma}^T \tilde{\mathbf{n}}_k)$. Here $\otimes$ is Kronecker operator. This way, we can solve for the two-dimensional feature matrix at the eigen-domain level similarly to (15) and (18), or

$$\boldsymbol{\beta} = \boldsymbol{\Xi}^{\mathbf{T}} (\boldsymbol{\Xi}\boldsymbol{\Xi}^{\mathbf{T}} + \gamma \mathbf{I})^{-1} \hat{\boldsymbol{\beta}}, \tag{28}$$

where γ is the regularization term. Thus, after we convert $\boldsymbol{\beta}$ back to matrix, we will obtain the desired super-resolution feature matrix.

5. Extensions to two-dimensional linear discriminant analysis of principal component matrix

Similarly to PCA, 2DPCA is more suitable for face representation than face recognition. For better performance in recognition task, LDA is necessary. Unfortunately, the linear transformation of 2DPCA reduces only the size of rows. However, if we apply LDA directly to 2DPCA, the number of the rows still equals to the height of original image. As a result, we are still facing the singular problem in LDA. Thus, a modified LDA, called *two-dimensional linear discriminant analysis* (2DLDA), based on the 2DPCA concept is proposed to overcome the SSS problem. Applying 2DLDA to 2DPCA not only can solve the SSS problem and the curse of dimensionality dilemma but also allows us to work directly on the image matrix in all projections. This way, the spatial structure information is still maintained. Moreover, the SSS problem has been remedy since the size of all scatter matrices cannot be greater than the width of face image. Our research group (Sanguangsat, 2006) are the first group that focus on the extension of discriminant analysis of principal component of Section 3.1 by two-dimensional projection, called *two-dimensional linear discriminant of principal component matrix*

6. Experimental results

Having assumed that we can perfectly obtain the information regarding to frame to frame motion, hence we can use these information to form the proper super-resolution matrix equation in (5). In our experiment settings, evaluation images were shifted by a uniform random integer, blurred with 4×4 Gaussian point spreading function with standard deviation 1, and downsampled by a factor of four to produce 16 low-resolution images for each high-resolution image. Using 9 (preselected) out of 16 complete set of frames of each image, we can construct the super-resolution subspaces and also super-resolution images, respectively. Our super-resolution subspace approach is then compared with pixel-domain super-resolution approach using the class-specific subspace for face and automatic target recognition. Here, we conduct and show experiments according to the algorithm proposed in Subsection 3.1 only. Ongoing experiments on the other reconstruction algorithms, i.e., *discriminant analysis of principal components, two-dimensional eigenface-domain based super-resolution, and 2DLDA of 2DPCA*, are conducting. Essentially, we expect very encouraging the recognition results.

6.1 Evaluation databases

Eigenface-domain super-resolution method is used as the baseline for comparison based on the well-known Yale and AR face databases (Yale, 1997; Martinez, 1998) and MSTAR non-face database (Center, 1997), respectively.

6.1.1 Yale Database

The Yale database contains 165 images of 15 subjects. There are 11images per subject, one for each of the following facial expressions or configurations: center-light, with glasses, happy,

left-light, without glasses, normal, right-light, sad, sleepy, surprised, and wink. All sample images of one person from the Yale database are shown in Fig. 1. Each image was manually cropped and resized to 100×80 pixels. In all experiments, the five image samples (centerlight, glasses, happy, leftlight, and noglasses) are used for training, and the six remaining images (normal, rightlight, sad, sleepy, surprise and wink) for test.

Fig. 1. The sample images of one subject in the Yale database

Fig. 2. The sample images of one subject in the AR database

6.1.2 AR database

The AR face database was created by Aleix Martinez and Robert Benavente in the Computer Vision Center (CVC) at the U.A.B. It contains over 4,000 color images corresponding to 126 people's faces (70 men and 56 women). Images feature frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf). The pictures were taken at the CVC under strictly controlled conditions. No restrictions on wear (clothes, glasses, etc.), make-up, hair style, etc. were imposed to participants. Each person participated in two sessions, separated by two weeks (14 days) time. The same pictures were taken in both sessions.

In our experiments, only 14 images without occlusions (sun glasses and scarf) are used for each subject, as shown in Fig. 2. All images were manually cropped and resized to 112×92 pixels, and then convert to 256 level gray scale images. The first five images per subject are used to train, and the remaining images to test.

6.1.3 MSTAR database

The MSTAR public release data set contains high resolution synthetic aperture radar data collected by the DARPA/Wright laboratory Moving and Stationary Target Acquisition and Recognition (MSTAR) program. The data set contains SAR images with size 128×128 of three difference types of military vehicles, i.e., BMP2 armored personal carriers (APCs), BTR70 APCs, and T72 tanks. The sample images from the MSTAR database are shown in Fig. 3. Because the MSTAR database is large, at this time, all images were centrally cropped to 32×32 pixels for evaluation purpose.

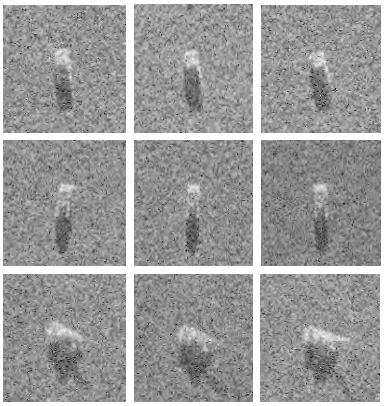


Fig. 3. Sample SAR images of MSTAR database: the upper row is BMP2 APCs, the middle row is BTR70 APCs, and the lower row is T72 tank.

Tables 1 and 2 detail the training and testing sets, where the depression angle means the look angle pointed at the target by the antenna beam at the side of the aircraft. Based on the different depression angles SAR images acquired at different times, the testing set can be used as a representative sample set of the SAR images of the targets for testing the recognition performance.

	Vehicle No.	Serial No.	Depression Angle	Images
BMP-2	1	9563	17^o	233
	2	9566		231
	3	C21		233
BTR-70	1	C71	17^o	233
T-72	1	132	17^o	232
	2	812		231
	3	S7		228

Table 1. MSTAR images comprising training set

	Vehicle No.	Serial No.	Depression Angle	Images
BMP-2	1	9563	15^o	195
	2	9566		196
	3	C21		196
BTR-70	1	C71	15^o	196
T-72	1	132	15^o	196
	2	812		195
	3	S7		191

Table 2. MSTAR images comprising testing set

6.2 Class-specific subspace results

The class-specific super-resolution images reconstructed for classification with pixel-domain and eigen-domain based approaches are shown in Fig. 4 and 5, respectively. The first images

in the first column are the input testing images. The images from the second to the sixth columns are corresponding to the class-specific super-resolution reconstruction obtained from the corresponding five different set of class-specific eigenfaces. Here, we show five class-specific units. Thus, five reconstructed images are obtained from each input image. Image with least error at i^{th} class-specific unit will be identified to i^{th} class. It should be noted that the images reconstructed using pixel-domain based super-resolution approach give us good perceptual view. However, as shown for eigen-domain based approach, the fourth and fifth input images also give us good perceptual views, while others give comparable reconstruction results. Thus, the reconstruction images based on class-specific super-resolution subspace are more dependent to its corresponding eigen-vectors.

Fig. 4. Samples of class-specific pixel-domain based super-resolution reconstruction images

Fig. 5. Samples of class-specific eigen-domain based super-resolution reconstruction images

	BMP-2	BTR-70	T-72	Recognition Acc.
BMP-2	526	0	61	89.61
BTR-70	103	0	93	0
T-72	19	0	563	96.74
Average	-	-	-	79.78

Table 3. The pixel-domain super-resolution based class-specific subspace method: Recognition test of a three class problem for 32 x 32 images.

	BMP-2	BTR-70	T-72	Recognition Acc.
BMP-2	526	0	61	89.61
BTR-70	116	0	80	0
T-72	41	0	541	92.96
Average	-	-	-	78.17

Table 4. Our proposed method: Recognition test of a three class problem for 32 x 32 images.

Database	Pixel-Domain	Eigen-Domain
Yale	88.56	81.11
AR	88.00	87.50
MSTAR	79.78	78.17

Table 5. Comparison Results for 32 x 32 images.

Table 3 and 4 show the confusion matrices of the MSTAR target recognition. As shown in Table 5, the performance of the pixel-domain based super-resolution method is slightly better than our proposed method. However, our method is greatly benefits in term of computation. Additionally, we can derive principal component coefficients of the face databases using simple matrix inversion of very small size, which is 36×36 only. This is because of the reason we use inner product approach to calculate the PCA coefficients. Thus, our algorithm is far faster than implementing super-resolution at pixel-domain. In pixel-domain based super-resolution approach, they have to solve a very large and sparse matrix using conjugate gradient method. In the MSTAR database, we found that the class 2 target cannot be recognized at all. This may be because the size of the low-resolution test image is too small. If we increase the size of the test images to 48×48 or larger, we think that we can have better recognition accuracy.

7. Conclusion

In this chapter we have conducted experiments on face and automatic target recognition by focusing on the eigenface-domain based super-resolution implementations. We have also presented an extensive literature survey on the subject of more advanced and/or discriminant eigenface subspaces. From our discussion, several new super-resolution reconstruction algorithms have been proposed here.

In particular, several new eigenface-domain super-resolution algorithms are suggested as follows

1. Class-specific face subspace based super-resolution is proposed in Subsection 3.1
2. Equation (18) is used for including discriminant analysis of principal components for extracting face feature for eigenface-domain super-resolution
3. Equation (28) is used for two-dimensional eigenface-domain super-resolution
4. Two-dimensional eigenface in Equation (28) is proposed to be replaced by two-dimensional linear discriminant analysis of principal component matrix

Current research in face and automatic target recognition is yet to utilize the full potential of these techniques. During preparing this chapter, we have just realized that there many aspects of studies and comparisons that should be conducted to gain more understanding on the variants of the eigenface-domain based super-resolution. For example, recognition accuracy should be compared between majority-voting using multiple low-resolution eigenfaces VS one super-resolved eigenface. This way, we can relate a set of LR face recognition with multiple classifier system. Furthermore, all of the proposed algorithms use a two-stage approach, that is, dimensionality reduction is first implemented, after that the super-resolution enhancement is performed. It may be a little more encouraging if we can further conduct the study on *joint dimensionality reduction-resolution enhancement.* This idea is quite similar to *joint source-channel coding,* which is a very popular approach studied for transmitting data over network. Evidently, we are thinking about computing certain desired eigenfaces and then super-resolve the computed eigenfaces on the fry. This approach trends to be quite a more biological plausible.

8. Acknowledgment

The MSTAR data sets provided through the Center for Imaging Science, John Hopkin University, under the contact ARO DAAH049510494.This work was partially supported by the Thailand Research Fund (TRF) under grant number: MRG 5080427.

9. References

Park, S. C.; Park, M. K. & Kang, M. G. (2003). Super-Resolution Image Reconstruction: A Technical Overview. IEEE Signal Processing Magazine, Vol. 20, pp. 21-36

Nguyen, N.; Milanfar, P. & Golub, G. H (2001). A Computationally Efficient Image Super-Resolution Algorithm. IEEE Trans. Image Proc., Vol. 4, pp. 573-583

Lin, F.; Cook, J.; Chandran, V. & Sridharan, S. (2005). Face Recognition from Super-Resolved Images, ISSPA, Australia, August 2005

Wagner, R.; Waagen, D. & Cassabaum, M. (2004). Image Super-Resolution for Improved Automatic Target Recognition, SPIE Defense & Security Symposium, Orlando, USA, April 2004

Gunturk, B. K.; Batur, A. U.; Altunbasak, Y.; Hayes, M. H. & Mersereau, R. M. (2003). Eigenface-Domain Super-Resolution for Face Recognition. IEEE Trans. Pattern Anal. and Mach. Intell., Vol. 12, pp. 597-606

Jia, K. & Gong, S. (2005). Multi-Modal Tensor Face for Simultaneous Super-Resolution and Recognition, Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV), pp. 1683-1690, Washington, DC, USA, 2005

Sezer, O.G.; Altunbasak, Y. & Ercil, A. (2006). Face Recognition with Independent Component-based Super-Resolution, Proc. of SPIE: Visual Communications and Image Processing, Vol. 6077, 2006

Shan, S.; Gao, W. & Zhao, D. (2003). Face Recognition based on Face-Specific Subspace. International Journal of Imaging Systems and Technology, Vol.13, No.1, pp.23-32

Belhumeur, P. N.; Hespanha, J. P. & Kriegman, D. J. (1997). Eigenface vs. Fisherfaces: Recognition using Class Specific Linear Projection, IEEE Trans. Pattern Anal. Machine Intell, Vol.19 (May 1997), pp.711-720

Zhao, W.; Chellappa, R. & Krishnaswamy, A. (1998) Discriminant Analysis of Principal Components for Face Recognition, Proceedings of the 3rd IEEE International Conference on Face and Gesture Recognition (FG), pp. 336-341, Nara, Japan, 14-16 April 1998

Yang, J.; Zhang, D.; Frangi, A.F. & Yang, J. yu (2004) Two-dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition. IEEE Trans. Pattern Anal. and Mach. Intell., Vol.26, pp.131-137, Jan. 2004.

Sanguansat, P.; Asdornwised, W.; Jitapunkul, S. & Marukatat, S. (2006).Two-Dimensional Linear Discriminant Analysis of Principle Component Vectors for Face Recognition. IEICE Transaction on Information and System: Special Issue on Machine Vision Applications, Vol. E89- D, No. 7, pp. 2164-2170

Vijay Kumar, B. G. & Aravind, R. (2008). A 2D Model for Face Superresolution. Proceedings of the 17th International Conference on Pattern Recognition (ICPR), pp. 1–4, Tampa, Florida, USA, 2008

Gsmooth (n.d.). Gaussian Smoothing. Available from http://homepages.inf.ed.ac.uk/rbf/HIPR2/gsmooth.htm

Schott, J. R. (2005), Matrix Analysis for Statistics, John Wiley & Sons, ISBN 0-471-66983-0, New Jersey, USA

Yale University (1997). The Yale Face Database. Available from http://cvc.yale.edu/projects/yalefaces/yalefaces.html.

Martinez, A. & Benavente, R. (1998). The AR Face Database. Available from http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html.

The Center for Imaging Science (CIS) (1997). DARPA Moving and Stationary Target Recognition Program (MSTAR). Available from http://cis.jhu.edu/data.sets/MSTAR.

Face and Automatic Target Recognition Based on Super-Resolution Class-Specific Subspace

Widhyakorn Asdornwised[a]

[a]Digital Signal Processing Research Laboratory
Department of Electrical Engineering,
Chulalongkorn University,
Bangkok, 10330, THAILAND

ABSTRACT

Recently, super-resolution reconstruction (SRR) method of low-dimensional face subspaces has been proposed for face recognition. However, the reconstructed features obtained from the face-specific super-resolution subspace contain no class information. This paper proposes a novel method for super-resolution reconstruction of class-specific features that aims on improving the discriminant power of the recognition systems. Our experimental results on Yale and ORL face databases are very encouraging. Furthermore, the performance of our proposed approach on the MSTAR database is also tested for preliminary evaluation.

Keywords: Super-resolution image reconstruction, Eigenface, Class-specific subspace, MSTAR

1. INTRODUCTION

In general, the fidelity of data, feature extraction, discriminant analysis, and classification rule are four basic elements in face and target recognition systems. One of the efficacy of recognition systems could be improved by enhancing the fidelity of the noisy, blurred, and undersampled images that are captured by the surveillance imagers. Regarding to the fidelity of data, when the resolution of the captured image is too small, the quality of the detail information becomes too limited, leading to severely poor decisions in most of the existing recognition systems. Having used super-resolution reconstruction algorithms,[1] it is fortunately to learn that a high-resolution (HR) image can be reconstructed from an undersampled image sequence obtained from the original scene with pixel displacements among images. This HR image is then used to input to the recognition system for improved performance. In fact, super-resolution can be considered as the numerical and regularization study of the ill-conditioned large scale problem given to describe the relationship between low-resolution (LR) and HR pixels.[2]

On the one hand, feature extraction aims at reducing the dimensionality of face or target image so that the extracted feature is as representative as possible. On the other hand, super-resolution aims at visually increasing the dimensionality of face or target image. Having applied super-resolution methods at pixel domain,[3,4] the performance of face and target recognition applicably increases. However, with the emphases on improving computational complexity and robustness to registration error and noise, the continuing research direction of face recognition is now focusing on using eigenface super-resolution.[5–7]

It should be noted that the above eigenface-domain super-resolution face recognition methods use the subspace projections that do not consider class label information. The eigenface's criterion chooses the face subspace (coordinates) as the function of data distribution that yields the maximum covariance of all sample data. In fact, the coordinates that maximize the scatter of the data from all training samples might not be so adequate to discriminate classes. In recognition task, a projection is always preferred to include discrimination information between classes. One of the extensions of eigenface, called *face-specific subspace* (FSS),[8] is proposed as an alternative feature extraction method to include class information for face recognition application. According to FSS, each reduced dimensional basis of *class-specific subspace* (CSS) is learned from the training samples of the same class. Actually, each individual set of CSS optimally represents the data within its own class with

Further author information: (Send correspondence to W.A.)
W.A.: E-mail: widhyakorn.a@chula.ac.th, Telephone: +66 (2) 218 6907

negligible error. As a result, large representation error occurs, when the input data is projected and then reconstructed using a reduced set with less maximum covariance coordinates (or equivalently, using a set of principal components that does not belong to the input class). This way, by using reconstruction error obtained from projection-reconstruction process between class, also called *distance from CSS* (DFCSS), a new metric can be suitably used as the distance for classifying the input data. In other words, the smaller the DFCSS is, the higher the probability that the input data belongs to the corresponding class will be.

There are many advantages of using CSS in face and target recognition. For example, the transformation matrices are trained from samples within their own classes, thus it is more optimum (using fewer components) to represent each sample in its own class than a transformation matrix trained by samples in all classes. Additionally, since DFCSS is the distance between the original image and its reconstruction image obtained from CSS, the memory space needed is only for storing the N transformation matrices, where N is the number of classes. This is far less than the conventional subspace methods, where we need to store both a single all-classes transformation matrix and also its prototypes (a large set of feature vectors calculated for all training samples). Moreover, the number of distance calculation in CSS is less than the number of distance calculation in conventional methods, since the number of classes is usually less than the number of training samples.

In this paper, we propose an efficient super-resolution method for face and target recognition that aims on including low-dimensional discriminant information within the super-resolution framework in order to improve performance. In Section 2, the idea of super-resolution for low-dimensional framework is formulated. Then, class-specific recognition approach is detailed in Section 3. By combining super-resolution reconstruction approach with class-specific idea, we propose a new method for face and automatic target recognition. Section 4 provides the experimental results for the Yale and ORL face databases and non-face database, MSTAR. Finally, conclusions are presented in Section 5.

2. SUPER-RESOLUTION FOR LOW-DIMENSIONAL FRAMEWORK

By arranging image pixels in lexicographical order according to the numerically computational SRR framework,[2] the relationship between an HR image and a set of LR images can be formulated in matrix form as follows:

$$\mathbf{f}_k = DCE_k\mathbf{x} + n_k, \qquad 1 \le k \le p \tag{1}$$

where D is the downsampling operator, C is the blurring or averaging operator, and E_k is the affine transform and n_k is noise of the frame k, respectively.

Thus, we can reformulated (1) as

$$\begin{bmatrix} \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_k \end{bmatrix} = \begin{bmatrix} H_1 \\ \vdots \\ H_k \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{n}_1 \\ \vdots \\ \mathbf{n}_k \end{bmatrix}, \tag{2}$$

or

$$\mathbf{f} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{3}$$

where

$$\mathbf{H} = \begin{bmatrix} DCE_1 \\ \vdots \\ DCE_k \end{bmatrix} = \begin{bmatrix} H_1 \\ \vdots \\ H_k \end{bmatrix}. \tag{4}$$

The above equation can be solved by adding a regularization term, or

$$\mathbf{x} = \mathbf{H^T}(\mathbf{H}\mathbf{H^T} + \lambda\mathbf{I})^{-1}\mathbf{f}. \tag{5}$$

It should be noted that matrix H is a large sparse matrix, which is very hard to find the solution of x One of the popular methods used to find the solution of the above inverse problem is the conjugate gradient method.

Dimensionality reduction of data is a common preprocessing step used for pattern recognition and classification applications and in compression schemes. In image analysis, principal component analysis (PCA) is also one of the popular methods used. Let $\mathbf{\Phi}$ be a reversible transformation that removes the redundancy by decorrelating the image data $\mathbf{x}$. Assuming face image $\mathbf{x}$ is ordered in lexicographic, thus, the represent ion form of $\mathbf{x}$ can be written as

$$\mathbf{x} = \mathbf{\Phi a} + \mathbf{e_x}, \tag{6}$$

where $\mathbf{a}$ is the $L \times 1$ dimensional feature that represents $\mathbf{x}$, and $\mathbf{e_x}$ is its representation error. Given that $\mathbf{\Psi}$ is $\beta^2 N^2 \times L$ dimensional reversible transformation matrix where β is the scaling resolution factor $(0 < \beta < 1)$, we can formulate low-resolution image representation as

$$\mathbf{f_k} = \mathbf{\Psi}\widehat{\mathbf{a}}_\mathbf{k} + \mathbf{e_{f_k}}. \tag{7}$$

By substituting (6) and (7) in (4), we obtain

$$\mathbf{\Psi}\widehat{\mathbf{a}}_\mathbf{k} + \mathbf{e_{f_k}} = \mathrm{H_k}\mathbf{\Phi a} + \mathrm{H_k}\mathbf{e_x} + \mathrm{n_k}. \tag{8}$$

Since

$$\mathbf{\Psi^T e_{f_k}} = \mathbf{0} \tag{9}$$

and

$$\mathbf{\Psi^T \Psi} = \mathbf{I}, \tag{10}$$

it is easy to derive the following equation

$$\widehat{\mathbf{a}}_\mathbf{k} = \mathbf{\Psi^T}\mathrm{H_k}\mathbf{\Phi a} + \mathbf{\Psi^T}\mathrm{H_k}\mathbf{e_x} + \mathbf{\Psi^T}\mathrm{n_k}. \tag{11}$$

Given

$$\zeta_\mathbf{k} = \mathrm{H_k}\mathbf{e_x} + \mathrm{n_k} \tag{12}$$

and

$$\Lambda_k = \mathbf{\Psi^T} H_k \mathbf{\Phi} \tag{13}$$

as the new observation noise with Gaussian distribution (see detail[5]) and the new linear transformation matrix respectively, we can obtain

$$\widehat{\mathbf{a}}_\mathbf{k} = \Lambda_\mathrm{k}\mathbf{a} + \mathbf{\Psi^T}\zeta_\mathbf{k}. \tag{14}$$

Without loss of generality, we can numerically solve for the true PCA feature vector as in (5), or

$$\mathbf{a} = \mathbf{\Lambda^T}(\mathbf{\Lambda\Lambda^T} + \lambda\mathbf{I})^{-\mathbf{1}}\widehat{\mathbf{a}}. \tag{15}$$

3. CLASS-SPECIFIC SUPER-RESOLUTION SUBSPACE

The original face-specific subspace (FSS) was proposed for applying to eigenface to improve performance. According to FSS, the difference between CSS and the traditional method is that the covariance matrix of the p^{th} class is individually evaluated from training samples of the p^{th} class. Thus, the p^{th} CSS is represented as a 4-tuple: the projection matrix, the mean of the p^{th} class, the eigenvalues of covariance matrix and the dimension of the p^{th} CSS. For identification, the input sample is projected using all CSSs and then reconstruct by those CSSs. If reconstruction error which obtained from the p^{th} CSS is minimum then the input sample is belong to the p^{th} class, also called distance from CSS (DFCSS).

By combining super-resolution reconstruction approach with class-specific idea, we propose a new method for face and automatic target recognition.

Figure 1. The sample images of one subject in the Yale database.

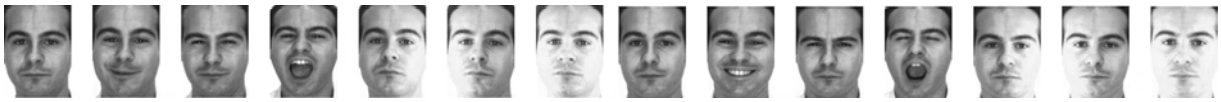


Figure 2. The sample images of one subject in the AR database.

4. EXPERIMENTAL RESULTS

Eigenface-domain super-resolution method is used as the baseline for comparison based on the well-known Yale,[9] AR[10] face database and MSTAR[11] non-face databases, respectively. The test images were shifted by a uniform random integer, blurred with 4×4 Gaussian point spreading function with standard deviation 1, and downsampled by a factor of four to produce 16 low-resolution images for each high-resolution image. Using 9 (preselected) out of 16 complete set of frames of each image, we can construct the super-resolution subspaces and also super-resolution images for face recognition evaluation.

Having assumed that we can perfectly obtain the information regarding to frame to frame motion, hence we can use these motions to form the proper super-resolution matrix equation in (5). Our super-resolution subspace approach is then compared with pixel-domain super-resolution approach using the class-specific subspace for face recognition.

4.1. Experiments on Yale database

The Yale database contains 165 images of 15 subjects. There are 11images per subject, one for each of the following facial expressions or configurations: center-light, with glasses, happy, left-light, without glasses, normal, right-light, sad, sleepy, surprised, and wink. All sample images of one person from the Yale database are shown in Fig. 1. Each image was manually cropped and resized to 100×80 pixels.

In all experiments, the five image samples (centerlight, glasses, happy, leftlight, and noglasses) are used for training, and the six remaining images (normal, rightlight, sad, sleepy, surprised and wink) for test.

4.2. Experiments on AR database

The AR face database was created by Aleix Martinez and Robert Benavente in the Computer Vision Center (CVC) at the U.A.B. It contains over 4,000 color images corresponding to 126 people's faces (70 men and 56 women). Images feature frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf). The pictures were taken at the CVC under strictly controlled conditions. No restrictions on wear (clothes, glasses, etc.), make-up, hair style, etc. were imposed to participants. Each person participated in two sessions, separated by two weeks (14 days) time. The same pictures were taken in both sessions.

In our experiments, only 14 images without occlusions (sun glasses and scarf) are used for each subject, as shown in Fig. 2. All images were manually cropped and resized to 112×92 pixels, and then convert to 256 level gray scale images. The first five images per subject are used to train, and the remaining images for test.

4.3. Experiments on MSTAR database

The MSTAR public release data set contains high resolution synthetic aperture radar data collected by the DARPA/Wright laboratory Moving and Stationary Target Acquisition and Recognition (MSTAR) program. The data set contains SAR images with size 128×128 of three difference types of military vehicles – BMP2 armored personal carriers (APCs), BTR70 APCs, and T72 tanks. The sample images from the MSTAR database

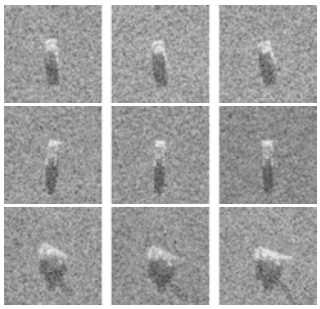


Figure 3. Sample SAR images of MSTAR database: the upper row is BMP2 APCs, the middle row is BTR70 APCs, and the lower row is T72 tank.

Table 1. MSTAR images comprising training set.

	Vehicle No.	Serial No.	Depression	Images
BMP-2	1	9563	17^0	233
	2	9566		231
	3	c21		233
BTR70	1	c71	17^0	233
T-72	1	132	17^0	232
	2	812		231
	3	s7		228

are shown in Fig. 3. Because the MSTAR database is large, at this time, all images were centrally cropped to 32×32 pixels for evaluation purpose.

Tables 1 and 2 detail the training and testing sets, where the depression angle means the look angle pointed at the target by the antenna beam at the side of the aircraft. Based on the different depression angles SAR images acquired at different times, the testing set can be used as a representative sample set of the SAR images of the targets for testing the recognition performance.

4.4. Face and Target Recognition Results

Fig. 4-5 are the reconstruction images used for class-specific classification method using pixel-domain and eigen-domain based reconstruction approaches. Each image is reconstructed from each class-specific's corresponding eigen-vectors. Here, we show five class-specific units. So five reconstructed images are obtained from each input image. Image with least error at i^{th} class-specific unit will be identified to i^{th} class. Note that images reconstructed using pixel-domain based approach at their corresponding class-specific units give us good perceptual

Table 2. MSTAR images comprising testing set.

	Vehicle No.	Serial No.	Depression	Images
BMP-2	1	9563	15^0	195
	2	9566		196
	3	c21		196
BTR70	1	c71	15^0	196
T-72	1	132	15^0	196
	2	812		195
	3	s7		191

Table 3. The pixel-domain super-resolution based class-specific subspace method: Recognition test of a three class problem for 32 x 32 images.

	BMP-2	BTR-70	T-72	Percent
BMP-2	526	0	61	89.61
BTR-70	103	0	93	0
T-72	19	0	563	96.74
Avg. -	-	-	-	79.78

Table 4. Our proposed method: Recognition test of a three class problem for 32 x 32 images.

	BMP-2	BTR-70	T-72	Percent
BMP-2	526	0	61	89.61
BTR-70	116	0	80	0
T-72	41	0	541	92.96
Avg. -	-	-	-	78.17

view. In eigen-domain based approach, the fourth and fifth input images give a very good perceptual view, while others give comparable reconstruction results. The reconstruction images based on class-specific super-resolution subspace are more dependent to its corresponding eigen-vectors.

Table 3-4 show the confusion matrices of the MSTAR target recognition. As shown in Table 5, the performance of the pixel-domain based super-resolution method is slightly better than our proposed method. However, our method is greatly benefit in term of computation. Additionally, we can derive principal component coefficients of the face databases using simple matrix inversion of size only 36×36. This is because we use inner product approach for calculating the PCA coefficients. In pixel-domain based super-resolution approach, however, we still have to solve a very large and sparse matrix using conjugate gradient method. Additionally, we increased the low-resolution testing images of the reference frame to the original resolution and then used them to evaluate with the training images at the original resolution. We got 85.56 and 77.00 percents for the recognition accuracy of the Yale and AR Face databases, respectively. As the result, our proposed method is better than the interpolation method for the AR face database, but not as good as the interpolation method for the Yale face database.

In the MSTAR database, class 2 of the target cannot be recognized at all. We think if we test images with a size of 48×48 or larger, we can recognize its class more correctly.

Table 5. Comparisons of the Recognition Accuracy (in percent) of PCA, PCA–SRR, CSS-SRR, and MD-CSS–SRR on Yale, AR, and MSTAR databases

Database Method	CSS-PCA	Pixel-Domain SR-CSS	Eigen-Domain SR-CSS
Yale	88.89	88.56	81.11
AR	88.00	88.00	87.50
MSTAR	77.44	79.78	78.17

Figure 4. Samples of reconstruction images used for class-specific classification approach. The first images in the first column are the input testing images. Images of the second to the sixth columns are corresponding to the class-specific principal component analysis reconstruction. Note that all the images here are reconstructed using the pixel-domain based reconstruction approach.

5. CONCLUSIONS

In this paper, we proposed an alternative method to eigen-domain super-resolution for face recognition method. Our proposed scheme was inspired from the frameworks of class-specific pattern analysis. One of the advantages of class-specific concept is that the feature vectors are no need to be stored, only Ntransformation matrices are need to be stored, where N is the number of classes. This is not the case in the conventional PCA method. Thus, the memory consumption is reduced significantly.

So far, we have not record and compare training and recognition time of all the methods. But we certainly believe that our proposed method can be benefit in term of computational complexity from the class-specific and eigen-domain based super-resolution approaches.

6. ACKNOWLEDGMENTS

The MSTAR data sets provided through the Center for Imaging Science, John Hopkin University, under the contact ARO DAAH049510494. The author would like to thank the Rajadapisek Sompoj Endowment Fund, Chulalongkorn University.

REFERENCES

1. S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, pp. 21–36, 2003.
2. N. Nguyen, P. Milanfar, and G. H. Golub, "A computationally efficient image superresolution algorithm," *IEEE Trans. Image Proc.*, vol. 4, pp. 573–583, 2001.
3. Frank Lin, James Cook, Vinod Chandran, and Sridha Sridharan, "Face recognition from super-resolved images," *ISSPA*, August 2005, Australia.

Figure 5. Samples of reconstruction images used for class-specific classification approach. The first images in the first column are the input testing images. Images of the second to the sixth columns are corresponding to the class-specific principal component analysis reconstruction. Note that all the images here are reconstructed using the eigen-domain based reconstruction approach.

4. R. Wagner, D. Waagen, and M. Cassabaum, "Image super-resolution for improved automatic target recognition," *SPIE Defense & Security Symposium*, April 2004, Orlando, USA.
5. Bahadir K. Gunturk, Aziz U. Batur, Yucel Altunbasak, Monson H. Hayes, and Russell M. Mersereau, "Eigenface-domain super-resolution for face recognition," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 12, pp. 597–606, May 2003.
6. Kui Jia and Shaogang Gong, "Multi-modal tensor face for simultaneous super-resolution and recognition," *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision*, pp. 1683–1690, 2005, Washington, DC, USA.
7. Osman G. Sezer, Yucel Altunbasak, and Aytul Ercil, "Face recognition with independent component-based super-resolution," *Proc. of SPIE : Visual Communications and Image Processing*, vol. 6077, 2006.
8. S. Shan, W. Gao, and D. Zhao, "Face recognition based on face-specific subspace," International Journal of Imaging Systems and Technology, vol.13, no.1, pp.23–32, 2003.
9. Yale University, "The Yale face database," 1997. Available from http://cvc.yale.edu/projects/yalefaces/yalefaces.html.
10. A. Martinez and R. Benavente, "The AR face database," 1998. Available from http://rvl1.ecn.purdue.edu/∼aleix/aleix_face_DB.html.
11. The Center for Imaging Science (CIS), "DARPA Moving and Stationary Target Recognition Program (MSTAR)," 1997. Available from http://cis.jhu.edu/data.sets/MSTAR.